

BUREAU OF EDUCATIONAL RESEARCH
COLLEGE OF EDUCATION

A Critical Summary of Research
Relating to the Teaching of
Arithmetic

By

WALTER S. MONROE

Director, Bureau of Educational Research

and

MAX D. ENGELHART

Assistant, Bureau of Educational Research


PUBLISHED BY THE UNIVERSITY OF ILLINOIS, URBANA

1931

PREFACE

In view of the fact that several summaries of research in the field of arithmetic are available, the preparation and publication of another one requires justification. As the title indicates, the present summary is restricted to research relating to methods of learning and teaching arithmetic. A more significant characteristic is the attempt to effect a systematic and critical evaluation of the researches summarized. There have been several assertions that a considerable portion of the research reported during recent years is faulty, and a few studies have been criticized by writers in educational periodicals. In the preparation of summaries of research, however, there has been very little evaluation of the studies included. Although the authors of this bulletin have recognized certain specified criteria in their evaluation, the judgments are largely subjective, and, consequently, the conclusions relative to the dependable findings concerning the teaching of arithmetic may not be entirely valid. It is hoped, however, that the publication of this bulletin will contribute to a more adequate understanding of what a critical summary involves.

Controlled experimentation has been hailed as a means of securing dependable evaluations of all factors of the teaching process. Careful study, however, indicates certain significant difficulties, and it is hoped that the discussion in the final chapter of this bulletin will contribute to a saner understanding of experimentation. The expenditures required for certain types of studies do not appear to represent wise investments, and those who are interested in educational research should give careful attention to the probable dependability of the outcomes of the studies they undertake or sponsor.



Digitized by the Internet Archive
in 2023

TABLE OF CONTENTS

	PAGE
CHAPTER I. INTRODUCTION	7
CHAPTER II. METHODS OF TEACHING AND LEARNING THE FUNDAMENTALS	14
CHAPTER III. DRILL IN THE FUNDAMENTALS	30
CHAPTER IV. METHODS OF TEACHING PUPILS TO SOLVE VERBAL PROBLEMS.	50
CHAPTER V. METHODS OF DIAGNOSIS AND REMEDIAL TREATMENT.	68
CHAPTER VI. METHODS OF TEACHING READING OF ARITHMET- ICAL SUBJECT-MATTER.	76
CHAPTER VII. MOTIVATION OF LEARNING IN ARITHMETIC. .	81
CHAPTER VIII. GENERAL SUMMARY AND CONCLUSION . . .	89
CHAPTER IX. BIBLIOGRAPHY OF INVESTIGATIONS PERTAINING TO METHODS OF TEACHING ARITHMETIC.	108

CHAPTER I

INTRODUCTION

General purpose of this bulletin. The general purpose of this bulletin is to present a summary and an evaluation of the research relating to instructional methods employed in teaching arithmetic in Grades I to VIII. For each group of investigations the discussion appears under three heads: (1) summary of reported conclusions, (2) evaluation of experiments, (3) justified conclusions.

Sources of references to investigations. The sources of practically all of the references were the "Summary of Educational Investigations Relating to Arithmetic" of Buswell and Judd¹ and the annual supplements prepared by Buswell.² An investigation of Brownell on the techniques employed in research on arithmetic was of service in locating in the above summaries investigations of the types desired.³ The writers were able to include in the present summary some references not given in the sources cited above.

General types of research included. Most of the investigations included in this summary may be characterized as experiments. Many of these experiments are of the single-group type, and as such may be labeled "uncontrolled" experiments. In investigations of this kind, the experimenter subjects a single group of pupils to the method or procedure which he wishes to try out, and estimates by observation or by administering tests the improvement in achievement assumed to be due to the new method or procedure. Where the gains in achievement are large, the new method may, with some justification, be claimed effective, but it is evident that usually an unknown amount of the gain is due to the operation of other factors. Investigations of this kind are termed "experiments," even though uncontrolled factors are operative, because they possess one important characteristic of all experimentation—that of trying something out to see what happens.

A number of the experiments referred to in this summary are of the controlled type. In place of a single group, two or more equivalent groups of pupils are used. In the typical controlled experiment,

¹Buswell, G. T. and Judd, C. H. "Summary of Educational Investigations Relating to Arithmetic," *Supplementary Educational Monographs* No. 27. Chicago: University of Chicago Press, 1925. 212 p.

²These supplements are published in the *Elementary School Journal*, as for example: Buswell, G. T. "Summary of Arithmetic Investigations, 1928," *Elementary School Journal*, 29:691-8, 737-47; May, June, 1929.

³Brownell, W. A. "The Techniques of Research Employed in Arithmetic," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 415-443.

the two groups of pupils are equated with respect to intelligence or achievement test scores, or both; hence, they are considered potentially equivalent with respect to the planned instruction. These groups are subjected to instruction differing only with respect to the experimental factor. For example, one of the groups is taught to add in the upward direction, while the other group is taught to add in the downward direction. After a period of such instruction, in which attempts are made to prevent irrelevant factors from operating unequally on the two groups, the final achievement test is given. The difference in final-test scores, or in mean gains in achievement from initial to final tests, is then computed, and interpretations are made with respect to the relative superiority of the one method or of the other.

Several laboratory experiments have been included in this summary. In these investigations, laboratory apparatus, such as that used in recording eye-movements, is used to secure an understanding of the characteristics of arithmetical learning activity. Some of the investigations are of the type in which data are collected by means of a single administration of a test. In a few places in this summary, relevant "case studies" are cited. Previous summaries of research in the field of arithmetic have occasionally been used to supplement the judgments of the writers with respect to the original research.

It should be mentioned that investigations of the nature of pupil responses, as for example, the researches on the relative difficulty of the number combinations, have been excluded from this study. The same is true of analyses of arithmetic texts and practice materials. Research of this kind, however important, is, in the judgment of the writers, more relevant to the problems of the arithmetical curriculum than to problems of methods of teaching arithmetic.

Criteria recognized in the evaluation of investigations. Evaluation of experiments is largely a subjective matter, but the utilization of specified criteria will tend to make it more dependable. A critical reader may apply these same criteria to the experiments evaluated in this summary and determine, to his own satisfaction at least, whether or not the evaluations of the present writers are justified.

1. Definition and restriction of the experimental factor. In experimental investigations of methods of teaching, the ideal procedure is to vary one of the factors that affect pupil achievement while all others are kept constant. The factor that is varied is designated as "experimental," and, obviously, it must be defined in specific terms. Otherwise the basis of the experimentation cannot be definitely known. For example, if the method of instruction is the experimental

factor and is designated merely as "the project method versus the traditional method," the precise nature of the variation is not clear. Usually the experimental factor must be restricted to a single phase or detail of method. If it is complex, the experimenter cannot know which element of the method produced the observed effect in the pupil achievement. Hence, the factor that is being made the basis of experimentation must be defined and restricted in such a way that the results may be interpreted in definite terms.

2. Control of pupil factors. Variation in the experimental factor is secured by employing two or more groups of pupils and maintaining a specified status of this factor for each of the groups. For example, if the type of drill exercises on addition of integers is the experimental factor, one type is used with Group A, a second type, with Group B, a third type, with Group C, and so on. Since achievement is influenced by the capacity of the pupils to learn, by their previous school experience, by their interest in the field of learning, and the like, it is obviously necessary that all significant pupil factors be controlled. This control is usually secured by forming groups that are equivalent with respect to all significant pupil factors. Hence, unless some other means of control is effected, the degree of equivalence of the groups is a criterion of the dependability of the results of the experiment.

3. Control of important non-experimental factors. The achievement of pupils is affected by several factors. The more important ones appear to be the following:

1. Instructional techniques
2. Skill of the teacher in using the instructional techniques
3. Zeal of the teacher
4. Personality traits of the teacher
5. Instructional materials
6. Time spent in learning activity

The significance of these factors varies with the character of the achievement, but usually none of them should be neglected. The skill and the zeal of the teacher appear to be more significant than is commonly realized. Control of these factors may be attained by securing equivalence or by determining the effect of variation and by making appropriate allowance for this effect in interpreting the results.

4. Accuracy and validity of measures of differences in achievement. An index of the relative effectiveness of two methods of instruction or of two types of instructional materials is obtained by computing the difference between the means of the scores on the test administered at the close of the experiment, or, preferably, between

the mean gains in achievement, obtained by subtracting the initial-test means from the final-test means. The obtained difference is affected by the variable and systematic errors of measurement. It is possible, if the coefficients of reliability of the tests used are known, to make appropriate allowances for variable errors of measurement. If the test is administered to both groups under approximately the same conditions, the possibly existing systematic errors of measurement, while they may raise or lower the means similarly, will not influence to a significant extent the difference of the means. It should be noted, also, that fluctuations of testing conditions tending to create systematic errors in certain groups of scores will tend to produce variable errors when several groups are combined. Hence, when the number of pupils is large, the systematic errors are likely to be less significant than when the group of pupils is small. It should be emphasized in this connection, however, that, when the groups of pupils and the obtained differences in achievement are relatively small, the systematic and variable errors of measurement are not likely to be of negligible significance. It is, therefore, essential that adequate recognition be given to their possible or probable influence. The probable effect of systematic errors cannot be calculated by any formula, and for this reason they are the more difficult to deal with.

The problem of an experiment usually specifies or implies the nature of the achievement on which the evaluation of the experimental factor is to be based. Hence, it is necessary to consider the extent to which the instruments used actually measure the specified or implied pupil achievements. This may not be the same as the usual validity of the test, because in this case one is concerned only with the extent to which the test measures the achievement designated in its specified or implied function. It is possible that a test may be more valid with respect to the instructional methods or materials of one group than of the other. For example, a test consisting of addition and subtraction examples in a mixed order would be more valid for a group that had had addition and subtraction taught together than it would be for a group that had had these processes taught separately. A test may also be valid with respect to the measurement of the more specific abilities engendered in arithmetic and yet be quite invalid with respect to such general outcomes as attitudes, ideals, and interests. If the achievement of one of the groups includes such outcomes, the differences in achievement obtained will contain errors of validity. The effect of invalidity is to introduce additional variable errors, and, as in the case of the variable errors of measurement, the effect tends to become negligible when the groups are large. However, the valid-

ity of the test used should not be neglected when interpreting smaller differences in gains.

5. Justification of generalization. If the preceding criteria have been satisfied, conclusion reported may be accepted as dependable with respect to the pupils participating in the experiment. If, however, the investigator wishes to generalize, his data must satisfy an additional criterion. They must be representative of the larger population to which the generalization is to be applied. If the sample of pupils used in the experiment was random, the investigator is justified in using the standard, or probable, error of sampling as an index of the representativeness of his groups. If, on the other hand, the sample was not random, the investigator must use other means to show the extent to which his sample is representative. While no specific rules may be stated, the investigator should consider all of the available evidence relative to the traits of the groups concerned. For example, if he has scores of his pupils on intelligence and standardized achievement tests, he may compare the means and standard deviations of these scores with the corresponding measures of the larger population. If this comparison indicates that his sample is typical of the larger population, generalizations may be accepted with a reasonable degree of confidence. If the data do not satisfy this criterion of representativeness, the investigator should refrain from generalizing, or limit his generalizations appropriately.

The application of these criteria. In the evaluation of the studies reviewed in this bulletin, the second and third criteria are most prominent. The reader, however, should not infer that the other criteria are not important. Usually the definition and restriction of the experimental factor are obvious, and the instructional techniques applicable in the teaching of arithmetic tend to be relatively specific rather than general. Hence, a large proportion of the experiments in the field being considered satisfy this criterion.

In the judgment of the dependability of the differences in achievement reported in the experiments summarized in this bulletin, some attention has been given to their "statistical" significance. The combined allowance to be made for variable errors of measurement and of sampling may be determined through the use of appropriate formulae.⁴ The employment of this procedure yields either the probable, or standard, error of the difference, and it is customary to recognize a difference as "statistically" significant when it is equal to, or greater than, 2.78 times its standard error or approximately 4.4 times its

⁴See pages 101 to 106.

probable error.⁵ When an obtained difference is 2.78 times its standard error, the chances are not less than 369 to 1 (interpreting the standard error as a limit) that the difference would have the same sign, or be in the same direction, as they would have been if variable errors of measurement and of sampling were eliminated. The "statistical" significance of a difference is, therefore, not very meaningful, since a difference may be "statistically" significant and yet be undependable because of other limitations of the data, such as lack of equivalence, failure to control non-experimental factors, variable errors of validity, and systematic errors of measurement, validity, and sampling. It is a safe assumption that any difference not "statistically" significant in the customary usage would not be of acceptable dependability if consideration is given to all of the probable faults of the data. On the other hand, if an obtained difference is "statistically" significant, its dependability is more certain because of this, but it is by no means guaranteed. In the estimation of the dependability of differences reported in the experiments reviewed in this summary, "statistical" significance has been recognized, therefore, as but one aspect of the matter.

The magnitude of possible systematic errors due to lack of equivalence, to failure to control non-experimental factors, to failure to secure comparable testing conditions in experimental and control groups, and to failure to measure the same outcomes in both groups is difficult to determine from the report of an experiment, unless the investigator explicitly refers to the matter.

Unless some unusual achievement is specified or implied, most tests designed to measure calculation skills are probably of rather high validity. They, of course, measure the current ability of pupils rather than the permanent residue of achievement. It is likely that the latter type of achievement should be considered, but few, if any, investigators have attempted to base their conclusions on it. Consequently, the present writers have not applied this more severe test in their evaluations. When the achievement to be measured includes abilities other than calculation skills, the validity of the measures is an important matter, but it is very difficult to determine the degree of validity.

The organization of the summary. This summary of research relating to instructional methods in arithmetic has been divided into six major divisions represented by the following rubrics: (1) methods

⁵Monroe, W. S. and Engelhart, M. D. "Experimental Research in Education," *University of Illinois Bulletin*, Vol. 27, No. 32, Bureau of Educational Research Bulletin No. 48. Urbana: University of Illinois, 1930, p. 59-76. See also:

McCall, W. A. *How to Measure in Education*. New York: Macmillan Company, 1922, p. 404-5.

of learning and teaching the fundamentals, (2) methods of drill in the fundamentals, (3) methods of teaching pupils to solve verbal problems, (4) methods of providing diagnosis and remedial treatment, (5) methods of teaching the reading of arithmetical subject-matter, (6) methods of motivating learning activity in arithmetic. A chapter is devoted to each of these divisions.

CHAPTER II

METHODS OF TEACHING AND LEARNING THE FUNDAMENTALS

The general nature of the experimental factor. The experimental factors of the studies summarized in this chapter are essentially methods of learning or performing the fundamental operations of arithmetic. Requesting pupils to add upward or to add downward, as the case may be, may be thought of as a method of teaching, but the essential element is the activity of the pupil. In the same way, requesting pupils to use the subtractive method of subtraction in which borrowing or decomposition is used, and directing pupils in the use of this method, may be regarded as a method of learning.

The research summarized in this chapter has been classified under the following heads: (1) addition; (2) subtraction; (3) division; (4) fractions, decimals, percentage, proportion, and denominate numbers.¹

ADDITION

1. Summary of conclusions as reported. The relative efficiency of upward and downward addition has been studied in one experiment and in two investigations of other types. In the experiment reported by Buckingham² the group that was taught to add downward attained greater, but not significantly greater, achievement in addition. From an analysis of test results, Cole³ reported that individuals add more accurately downward but less rapidly. Buckingham⁴ has also reported the findings of a questionnaire study in which it was discovered that while more people prefer to add upward when the column is long, they add downward when the column is short. On the basis of the logical advantages that he claims for downward addition, and because of this variation, Buckingham recommends that downward addition be taught.

Procedures for adding a column of figures have been studied in four experiments and in one investigation where an observation technique was used. Overman⁵ investigated the relative effectiveness

¹The absence of multiplication from this classification should be noted. The present writers have been unable to discover any experimental investigations of methods of teaching or learning multiplication.

²Buckingham, B. R. "Upward versus Downward Addition," *Journal of Educational Research*, 16:315-22, December, 1927. (18)

³Cole, L. W. "Adding Upward and Downward," *Journal of Educational Psychology*, 3:83-94, February, 1912. (29)

⁴Buckingham, B. R. "Adding Up or Down: A Discussion," *Journal of Educational Research*, 12:251-61, November, 1925. (15)

⁵Overman, J. R. "An Experimental Study of the Effect of the Method of Instruction on Transfer of Training in Arithmetic," *Elementary School Journal*, 31:183-90, November, 1930. (97)

of the following methods of teaching addition (and subtraction) of two- and three-place numbers in terms of transfer to untaught types, such as addition of four two-place numbers, two three-place numbers, one three-place number, and one one-place number.

(1) In Method A the pupils were shown how to perform the process, and there was no generalization or consideration of underlying principles (2) In Method B (generalization) the pupils were helped to formulate general methods of procedure from the specific types taught, and these generalizations were constantly emphasized throughout the teaching (3) In Method C (rationalization) the reasons and principles underlying the specific types taught were discussed with the pupils. The formulation of general rules of procedure was avoided as much as possible (4) In Method D (generalization and rationalization) general methods of procedure were formulated, and the underlying principles were discussed.

Method B was reported as the most effective, Method D was found to be almost as effective as Method A, and Method C, only slightly more effective than Method A, the least effective of all. In connection with his experiment on transfer of learning in addition and subtraction Olander⁶ investigated the effectiveness of instruction in generalizing groups of combinations. "For example, these children were led to recognize the law common to zero combinations. They noted that combinations appeared in reverse form such as $6 + 7$ and $7 + 6$, and they observed that a combination such as $10 - 6$ was intimately related to $6 + 4$." In his conclusions the investigator states that short daily instruction of this character had no significant effect on the arithmetic scores of the pupils taught by the method.

Conard and Arps⁷ discovered that strikingly superior results were secured by teaching children to "think results only."⁸ Ballenger⁹ concluded that it is effective to teach children, who have been having difficulty with addition, to break long columns into two parts and to add each part separately. Arnett¹⁰ reported that the most rapid and accurate individuals add the digits in regular serial order. Excessive combination, or rearrangement, of digits is detrimental to rate and accuracy, but a moderate amount proves beneficial to some individuals. Finally, Clark and Vincent¹¹ found that teaching the pupils to check their answers results in greater, but not significantly greater, accuracy.

⁶Olander, H. T. "Transfer of Learning in Simple Addition and Subtraction," *Elementary School Journal*, 31:358-69, 427-37; January, February, 1931. (94)

⁷Conard, H. E. and Arps, G. F. "An Experimental Study of Economical Learning," *American Journal of Psychology*, 27:507-29, October, 1916. (32)

⁸In this method the individual in the process of adding 3, 4, 9, and 6 thinks 7, 16, and 22 rather than 3 and 4 are 7, 7 and 9 are 16, and 16 and 6 are 22.

⁹Ballenger, H. L. "Overcoming Some Addition Difficulties," *Journal of Educational Research*, 13:111-17, February, 1926. (6)

¹⁰Arnett, L. D. "Counting and Adding," *American Journal of Psychology*, 16:327-36, July, 1905. (4)

¹¹Clark, J. R. and Vincent, E. L. "A Study of the Effect of Checking Upon Accuracy in Addition," *Mathematics Teacher*, 19:65-71, February, 1926. (27)

2. **Evaluation of the experiments.** In the only experiment on upward versus downward addition, Buckingham (18) used seven pairs of groups of second- and third-grade pupils, varying in size from eleven to twenty-eight pupils. The paired groups were equated with respect to scores made on an initial test in addition. Each of the teachers participating in this experiment taught a pair of groups, rotated at the end of each week the time of day during which addition was taught, assigned no home work in arithmetic, and introduced no new arithmetic topics. The teacher administered the final test as soon as her pair of groups had attained reasonable proficiency in adding short columns of one-place numbers. The differences in mean gains for the six pairs of groups favored the method of downward addition, but in only one case was the difference "statistically"¹² significant when compared with its probable error.

In this investigation, the experimental factor, the direction of adding a column of figures, is specific and appears to have been satisfactorily isolated. The control of the pupil factors by grouping the pupils on the basis of the scores made on an initial addition test probably was not entirely satisfactory. The general intelligence and the addition habits of the pupils were not directly considered. The control of the teacher factors was attempted by having each teacher instruct a pair of groups, one in adding upward and the other in adding downward. This procedure, however, does not insure control, because there may have been variations in zeal and skill. The validity of the test was not explicitly considered; it depends upon the ability that is specified as the criterion of merit of the direction of adding. If validity is defined as "ability to add throughout the pupils' school experience" or "ability to add when he becomes an adult," it must be admitted that the degree of validity is unknown. In view of the relatively small differences in gains, it seems reasonable to say that the findings, which are interpreted as favoring downward addition, are not dependable. When one considers the information yielded by Buckingham's questionnaire study (15) and by Cole's experiment with adults (29), it still appears that the relative merit of the two directions of adding has not been determined. Neither does one have dependable evidence to support the common-sense view that the direction makes little or no difference.¹³

Cole (29) had thirty persons add the same problems both upward and downward. The fact that the subjects were accustomed to use

¹²See pages 11 and 12.

¹³It may be somewhat immaterial whether children are taught to add upward or downward, since the carefully controlled experiment of Beito and Brueckner (9) would seem to indicate that there is a large amount of transfer of training from learning to add in one direction to learning to add in the other, or reverse, direction. It is stated in their conclusions that "When pupils of any mental level are taught

the upward method causes one to question the results obtained in this investigation. It is possible that they added downward more accurately because they added more slowly and took greater pains with an unfamiliar method.

Overman (97) used four groups of 112 second-grade pupils which were equivalent with respect to sex, mental age, teacher's estimate of general ability, and score on a preliminary test. The experimental factors¹⁴ appear to have been adequately defined, but there is some uncertainty in regard to the control of the non-experimental factors, especially teacher skill and zeal. Each of the groups were given twenty minutes of practice a day for fifteen days, eight days being used for testing, and seven, for instruction and practice. Tests were given at the beginning and at the end of the experiment, and twice during the experiment. The differences in achievement, as measured by these tests, were "statistically" significant for Methods B and D compared with Method A, but not for Method C. The conclusions of the experiment seem reasonably dependable. They also, for the most part, seem to be the conclusions one should logically expect. That pupils should be stimulated to generalize is sufficiently well established that an experimental comparison of a method with generalization and a method without generalization seems somewhat futile. One wonders in the case of this experiment why generalization plus rationalization should have proved inferior to generalization alone. Common sense would lead to the inference that a combination of both would be most effective. It would seem justifiable to ascribe the apparent inferiority not to the method which combines generalization with rationalization but to the limitations of the experiment.

In evaluating the effectiveness of instruction in generalizing groups of combinations, Olander (94) used three hundred pairs of second-grade pupils equivalent with respect to growth in arithmetic ability over a period of five weeks. The reason given for using this technique is the following:

If two groups exhibit similar learning curves under similar instruction until a certain point is reached, it can be assumed that the groups are equal in the function in question.

The experiment was conducted for twelve more weeks, during

only the direct form of an addition combination, such as $\frac{7}{4}$, as nearly as can be, the reverse form, $\frac{4}{7}$.

is learned concomitantly at least as completely as the direct form." See:

Beito, E. A. and Brueckner, L. J. "A Measurement of Transfer in the Learning of Number Combinations," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 569-87. (9)

¹⁴See page 15 for a description of these factors.

which time the pupils of the experimental group were given instruction in generalizing for three minutes of the daily twenty-minute period. Achievement was tested at the start of the experiment, at the end of five weeks, at the end of eleven weeks, and at the close of the experiment—at the end of seventeen weeks. The tests included the one hundred addition and the one hundred subtraction combinations and were administered by the flash-card method. The pupils in the group not given the three minutes of daily generalizing instruction were able to generalize practically as well as the pupils who were given this instruction. It appears logical that the experimental factor, the generalization instruction, was not applied long or intensively enough to add materially to the generalizing abilities acquired by the pupils on their own account. The interpretation that generalization instruction is not worth while, on the basis of Olander's data, does not seem to be justified.¹⁵

In studying the effect of teaching pupils to "think results only," Conard and Arps (32) used two groups of thirty-two grade-school children whose approximate equivalence was shown by comparison of scores on the Courtis test. After eight work periods of seven examples in each of the four fundamentals, the final test was administered. This experiment does not appear to justify a very high rating with reference to any of the criteria. The experimental factor was not adequately defined, and the control of the non-experimental factors probably was not sufficient to justify acceptance of the obtained results as demonstrating the superiority of "thinking results only." There is evidence that experimental conditions were in some respects abnormal and that the experimental pupils sometimes forgot to "think results only." It may be argued that these faults, for the most part, were such as would tend to reduce, rather than to increase, the difference in favor of the experimental method and, consequently, that the findings should be accepted as dependable evidence. In view of the limitations, however, this argument is not convincing, and the reported conclusion probably should not be accepted as dependable.

Ballenger (6) used a single group of 130 fourth-, fifth-, and sixth-grade children. These children were taught to divide long columns of figures and to add each part separately. While they improved significantly in accuracy, the results of this uncontrolled experiment cannot be regarded as other than merely suggestive. Such a procedure might be effective for backward children; it probably should not be recommended as a standard method of teaching addition. Children should be taught to add columns of increasing length.

¹⁵The other conclusions stated in this experiment appear to be reasonably dependable.

Splitting columns, as advocated by Ballenger, would seem to be a method of forming undesirable habits which would need to be unlearned later.

Arnett (4) used chronoscopic apparatus in determining the methods of counting and of adding used by several adults in a psychological laboratory. His results are suggestive, but they should be verified by observation and by controlled experimentation with school children.

Clark and Vincent (27), in their study of the effect of checking on accuracy, used two groups of fifth- and sixth-grade children which were equated on the basis of M. A., I. Q., and initial addition test scores. The size of these groups is not reported. After twenty days of practice, the final test was administered. The principal limitations of this experiment are to be found in the lack of control of non-experimental factors, in the lack of control of special teacher zeal, and in the unknown validity of the tests. The difference in final-test means was in favor of the method of checking, but not significantly so. This might be interpreted to mean that teaching pupils to check additions may be expected to increase the accuracy of their work only very slightly. This conclusion, however, probably is not justified.

3. Justified conclusions. It is evident that none of the experiments satisfy completely the criteria stated in Chapter I. Those of Buckingham (18), Overman (97), and Olander (94) come nearest to doing so, but the limitations of these experiments render the conclusions of somewhat doubtful dependability. More experiments must be reported before justified conclusions can be expressed with respect to such problems as adding upward versus adding downward, the effect of checking on accuracy, and the like. The merits of instruction involving generalization and rationalization should be tested in experiments where failure to control important non-experimental factors does not obscure the effectiveness of such instruction.

SUBTRACTION

The relative merits of the four principal methods of subtraction have been studied in a number of investigations. These methods may be described briefly by noting the steps in subtracting 25 from 43.

In using the *subtractive*, or *take-away*, method in which *borrowing* or *decomposition* is employed, the steps are:

$$5 \text{ from } 13 = 8$$

$$2 \text{ from } 3 = 1$$

In the *subtractive*, or *take-away*, method in which *carrying* or *equal addition* is used, the steps are:

$$5 \text{ from } 13 = 8$$

$$3 \text{ from } 4 = 1$$

The *additive* method in which *borrowing* or *decomposition* is used requires the following steps:

5 and what are 13, write 8

2 and what are 3, write 1

The *additive* method in which *carrying* or *equal addition* is used is illustrated as follows:

5 and what are 13, write 8

3 and what are 4, write 1

Decomposition, usually when used as illustrated in the first of these examples, has been called the "first Italian method," and equal addition, when used as in the second example, has been called the "second Italian method." No name is given to the third method, but the fourth is well known as the "Austrian method." Irmina¹⁶ has described a "complementary method" in which either decomposition or equal addition may be used. However, since no experimental evidence has been presented with respect to its merits, this method is not considered here.

1. **Summary of reported conclusions.** The conclusions of Buckingham,¹⁷ Mead and Sears,¹⁸ and Taylor¹⁹ favor the subtractive methods in comparison with the additive methods.²⁰ The only conclusion favorable to the additive methods is that of Beatty²¹ who found that greater accuracy but less speed resulted from their use. Ballard,²² McClelland,²³ and Winch,²⁴ studied the relative merits of decomposition, or borrowing, versus equal addition, or carrying, in connection with the subtractive procedure.²⁵ In each case the results favored the equal addition, or carrying, process.²⁶ Johnston's²⁷ pupils used both the subtractive and additive general methods. For

¹⁶Irmina, Sister M. "The Relative Merits of the Methods of Subtraction," *Catholic University of America, Educational Research Bulletins*, Vol. III, No. 9. Washington: Catholic Education Press, 1928, p. 4-5.

¹⁷Buckingham, B. R. "The Additive versus the Take-Away Method of Teaching the Subtraction Facts," *Educational Research Bulletin* (Ohio State University), 6:265-69, September 28, 1927. (16)

¹⁸Mead, C. D. and Sears, Isabel. "Additive Subtraction and Multiplicative Division Tested," *Journal of Educational Psychology*, 7:261-70, May, 1916. (72)

¹⁹Taylor, J. S. "Subtraction by the Addition Process," *Elementary School Journal*, 20:203-7, November, 1919. (114)

²⁰In other words, the conclusions favor the first two procedures illustrated on pages 19 and 20 rather than the last two.

The following study, not accessible to the writers, also favored the subtractive equal additions method.

²¹Methods of Subtraction," *St. Louis Public School Messenger*, 26:28-32, September 1, 1928. (128)

²²Beatty, W. W. "The Additive versus the Borrowing Method of Subtraction," *Elementary School Journal*, 21:198-200, November, 1920. (8)

²³Ballard, P. B. "Norms of Performance in the Fundamental Processes of Arithmetic, with Suggestions for Their Improvement," *Journal of Experimental Pedagogy*, 2:396-405, December 5, 1914; 3:9-20, March 5, 1915. (5)

²⁴McClelland, W. W. "An Experimental Study of the Different Methods of Subtraction," *Journal of Experimental Pedagogy*, 4:293-99, December 5, 1918. (69)

²⁵Winch, W. H. "Equal Additions versus 'Decomposition' in Teaching Subtraction: An Experimental Research," *Journal of Experimental Pedagogy*, 5:207-20, 261-70; June 5, December 6, 1920. (125)

²⁶The first two procedures illustrated on page 19.

²⁷The second of the procedures illustrated on page 19.

²⁸Johnston, J. T. "The Merits of Different Methods of Subtraction," *Journal of Educational Research*, 10:279-90, November, 1924. (52)

both of these groups equal addition, or carrying, was found to be superior.

2. Evaluation of experiments. McClelland (69), Mead and Sears (72), Winch (125), and Buckingham (16) experimented with school children. In all cases the experimental factor was defined and sufficiently restricted. The other criteria, however, were not fully satisfied. McClelland (69) employed two groups of children between twelve and one-half to thirteen and one-half years of age in an English school. One group of thirty-four had been accustomed to use the method of equal addition, and the other group of thirty-two, the method of decomposition. After an initial program of testing, which revealed that the equal-addition group was significantly superior, the groups were practiced in their respective methods for a period of twenty weeks. The equal-addition group achieved the greater per cent increase in speed and accuracy. It is evident that McClelland is to be criticized for failure to secure equivalence at the beginning of his experiment. It is possible that the group using the equal-addition method consisted of more intelligent children and, in consequence, made the greater gain in achievement. Furthermore, the degree of control of non-experimental factors is not known.

Winch (125) conducted two experiments with girls in English schools. In the first, two groups of nineteen eleven-year-old girls were equated on the basis of scores on a series of initial subtraction tests. All of the children had previously used the decomposition method. In the experiment, one group was practiced in this method, while the other learned the equal-addition method. After eight lessons of fifteen to twenty minutes each the achievement of the group learning the equal-addition method slightly surpassed that of the other, as shown by the scores on a series of final subtraction tests. The second experiment was conducted with two groups of twenty-three eight and one-half year old girls who had been accustomed to the equal-addition method. After equivalence had been secured with respect to ability to subtract, one group was practiced in the equal-addition method, while the other group learned the method of decomposition. After eight lessons of thirty minutes each, four final tests were given. The difference between the final-test means in favor of the equal-addition method is approximately seven times its probable error. Winch is to be commended for his care in securing equivalence with respect to initial subtraction ability, for efforts to control non-experimental factors, and for the statistical treatment of his results. He is to be criticized for the non-representativeness and smallness of his groups and for the short duration of his experiments. While the

techniques used in these experiments are in many respects excellent, it seems unsafe to generalize the findings reported.

Mead and Sears (72) used two second-grade classes of unreported size which were shown to be approximately equivalent with respect to ability in addition. One group was taught additive subtraction for four months while the other group learned the subtractive method. The final test revealed a possibly significant difference in favor of the subtractive method, so far as single-column subtraction was concerned. An additional test of three-figure-subtraction examples revealed no significant difference between the groups. Mead and Sears are to be criticized for failure to secure more adequate equivalence and for not revealing the size of their groups. They are to be commended for certain precautions taken to secure control of non-experimental factors and for their rather satisfactory interpretation of results.

In the experiment of Buckingham (16) seven pairs of groups ranging in size from five to twenty-nine pupils were equated in seven schools by means of the Pressey Primary Classification Test. Each of the teachers participating in the experiment taught both groups of a pair for a period of seven months, at the end of which time the pupils were tested for their proficiency in single-column subtraction. The differences in achievement for six of the seven groups favored the subtractive method as compared with the additive, but in no case was the difference "statistically" significant. Buckingham is to be commended for his techniques in securing equivalence, for using children of no initial ability in subtraction, and for certain precautions taken to secure control of non-experimental factors. He is also to be commended for using so many different groups and schools. The interpretation of his data would seem to exaggerate the effectiveness of the subtractive method. A more conservative interpretation would seem to be required.

Ballard (5), Beatty (8), Taylor (114), and Johnston (52) have reported the results of investigations in which the data were collected by test from pupils whose method of subtracting had been determined. Ballard (5) administered his test to 18,678 eight- and nine-year-old English school children. He found the achievement in subtraction in schools where equal addition, or carrying, was taught to be significantly superior to the achievement in schools where decomposition, or borrowing, was taught. He is to be criticized for failure to determine more adequately the methods actually used by the pupils. Taylor (114) had teachers of 11,368 fourth-, fifth- and sixth-grade children put a subtraction example on the board and determine, by

asking the children what they would say in solving the given example, the methods of subtraction that the children were using. His data showed that only 37.6 per cent were continuing to use the additive equal-addition method which they were supposedly taught, while the balance of the pupils had somehow learned and were using subtractive methods. Beatty (8) has criticized Taylor (114) for concluding that his results showed the inferiority of the additive equal-addition method, since evidence was not secured to prove that no other method was taught.

Beatty (8) administered the Courtis Research Standard Tests, Series B, to 54 pupils who used the additive methods and 115 pupils who used the borrowing (subtractive?) methods. While his results favor the additive methods for accuracy, they favor the borrowing methods for speed. He is to be criticized for his few cases and for failure to define the methods evaluated. He does contribute the information that 51.8 per cent of one group of eighty-three children actually did abandon the additive for borrowing methods.

Johnston (52) determined the subtraction methods used by 277 normal-school students and tested the students for speed and accuracy. His results are slightly significant with respect to the superiority of equal addition, or carrying, when used both with additive and subtractive methods, but are entirely inconclusive with respect to the additive versus subtractive methods. Ruch, Knight, and Lutes²⁸ have criticized Johnston for failure to make adequate allowance for the statistical limitations of his data. A computation by them of probable errors of the differences showed that none of the differences were "statistically" significant. Johnston²⁹ replied that their computations failed to consider the significant difference in speed in favor of the equal-addition method. When the accuracy means are corrected for speed, Johnston claims the difference is significant. Ruch, Knight, and Lutes³⁰ replied to this that no differences can be considered "statistically" significant from groups of eight, thirteen, or twenty-three cases. They add that the original report should have contained adequate information with respect to standard deviations and probable errors.

3. Justified conclusions. The great majority of the investigations favored the subtractive, or take-away, methods rather than the additive methods, and the equal-addition, or carrying, process rather than

²⁸Ruch, G. M., Knight, F. B., and Lutes, O. S. "On the Relative Merit of Subtraction Methods: Another View," *Journal of Educational Research*, 11:154-55, February, 1925.

²⁹Johnston, J. T. "Still on the Relative Merits of Subtraction Methods," *Journal of Educational Research*, 12:80-83, June, 1925.

³⁰Ruch, G. M., Knight, F. B., and Lutes, O. S. "A Rejoinder to Professor Johnston's Criticisms," *Journal of Educational Research*, 12:83-85, June, 1925.

that of decomposition, or borrowing. However, the faulty techniques used in these investigations, plus the failure to find truly significant differences in achievement between the different methods, would cause one to question the dependability of a conclusion in favor of the subtractive method in which equal addition is used, although the evidence is in its favor.

In this connection it is interesting to note that in two summaries of research in the field of arithmetic, Buswell favors the subtractive method in which equal addition is used.³¹ This conclusion agrees with that of Irmina,³² but differs with that of Knight, Ruch, and Lutes,³³ who present certain theoretical considerations in favor of the subtractive method in which borrowing or decomposition is used. Osburn³⁴ has reported a summary in which he computed the statistical errors of the differences given in the experimental literature. He states that the differences are significantly in favor of the subtractive equal-addition method as compared with the subtractive decomposition method, but the subtractive equal-addition method has not been shown to be significantly superior to the additive methods, although the chances are 16 to 1 in its favor. In another recent review of the subtraction experiments the opinion is expressed that "the differences among the rival methods of subtraction must be small; otherwise centuries of observation and a dozen empirical studies would long since have laid down the broad outlines of truth."³⁵

DIVISION

1. Summary of reported conclusions. There have been only two investigations of the methods of teaching and of learning division. Mead and Sears³⁶ report that multiplicative division is superior to the traditional method. They illustrate multiplicative division as follows:

The . . . multiplicative-division class said:
$$\begin{array}{r} 4 \\ 5 \overline{) 20} \end{array}$$
 20, five times what are twenty? Five times four are twenty.

Conard and Arps (32) reported that in division the most effective results are secured when pupils are taught to "think results only."

³¹Buswell, G. T. and Judd, C. H. "Summary of Educational Investigations Relating to Arithmetic," *Supplementary Educational Monographs*, No. 27. Chicago: University of Chicago Press, 1925, p. 78.

Buswell, G. T. "A Critical Survey of Previous Research in Arithmetic," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 460-61.

³²Irmina, *op. cit.*, p. 26-27.

³³Knight, F. B., Ruch, G. M., and Lutes, O. L. "How Shall Subtraction Be Taught?" *Journal of Educational Research*, 11:168, March, 1925.

³⁴Osburn, W. J. "How Shall We Subtract?" *Journal of Educational Research*, 16:237-46, November, 1927.

³⁵Ruch, G. M. and Mead, C. D. "A Review of Experiments on Subtraction," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 678.

³⁶Mead and Sears, *op. cit.*

2. **Evaluation of experiments.** Two third-grade classes of unreported size participated in the experiment by Mead and Sears (72). The initial test, which was in addition, showed some lack of equivalence so far as the trait tested was concerned. No other attempt was made to estimate the degree of equivalence. The division practice of both groups was restricted to simple division by fives. At the end of four months a possibly significant difference was found in favor of "multiplicative" division, as restricted in the preceding statement. A final test containing longer examples showed no significant difference between the groups. Mead and Sears are to be criticized for failure to secure equivalent groups, for failure to report the size of the groups used, for the restricted character of the training, and for attempting to correct for lack of equivalence in an unjustifiable manner. The units and zero points of the initial and final tests were shown to be different, and, therefore, correction by subtracting the difference between initial-test means from the difference between final-test means cannot be condoned.³⁷ Furthermore, there was not adequate control of the non-experimental factors.

The experiment of Conard and Arps (32) was evaluated under addition.³⁸

3. **Justified conclusions.** The faults of these two experiments make the listing of a justifiable conclusion impossible. It is doubtful whether the conclusion of Mead and Sears (72) should be regarded as indicative or suggestive.

FRACTIONS, DECIMALS, PERCENTAGE, AND PROPORTION, AND DENOMINATE NUMBERS

1. **Summary of reported conclusions.** Collier³⁹ has reported that children learn to multiply fractions effectively, if addition of fractions is used as a point of departure. For example, a child may be taught to multiply 4 by $\frac{2}{3}$ through a request to add $\frac{2}{3}$, $\frac{2}{3}$, $\frac{2}{3}$, $\frac{2}{3}$. When the result $\frac{8}{3}$ has been obtained by the child, the teacher should point out that 8 is the product of 4×2 . Anspaugh⁴⁰ has reported that drill on the fundamental combinations is effective in securing greater efficiency in handling common and decimal fractions.⁴¹

³⁷Monroe, W. S. and Engelhart, M. D. "Experimental Research in Education," *University of Illinois Bulletin*, Vol. 27, No. 32, Bureau of Educational Research Bulletin No. 48. Urbana: University of Illinois, 1930, p. 63. (Footnote 14)

³⁸See page 18.

³⁹Collier, Myrtle. "Learning to Multiply Fractions," *School Science and Mathematics*, 22:324-29, April, 1922. (30)

⁴⁰Anspaugh, G. E. "Teaching the Number Facts in the Komensky School," *Chicago Principals' Club, Second Yearbook*. Chicago: Chicago Principals' Club, 1927, p. 88-89. (2)

⁴¹Knight and Setzafandt have shown that training in the addition of fractions having certain denominators transfers to the addition of fractions having other denominators. Some inferences might be drawn from their conclusions with respect to effective methods of teaching the addition of fractions. See:

Knight, F. B. and Setzafandt, A. O. H. "Transfer within a Narrow Mental Function," *Elementary School Journal*, 24:780-87, June, 1924. (62)

Clapp, Chase, and Merriman⁴² found that practice material so prepared that it focuses the attention of the pupils on the kind of percentage problem they are attempting to solve is more effective than the ordinary textbook material. In ordinary textbook material problems solved similarly are grouped together, but in the experimental material "the pupil is not aided in solving the second problem (of a group of problems) by having solved the first one, unless he begins to understand the principle that underlies the solution of such problems." The nature of the problem statements is varied in the experimental material, and some problems not involving percentage are included to keep the minds of the pupils alert to the kinds of problems they are solving.

Monroe⁴³ concluded that children do not learn to place the decimal point in a quotient by a general rule, or as the result of the acquisition of a general ability. He contends that the placing of the decimal point in quotients requires several specific abilities.

Drushel⁴⁴ investigated the relative merits of two methods of placing the decimal point in long division by a test administered to college freshmen. In Method A the student used the rule: "There are as many places in the quotient as those in the dividend exceed the divisor." In Method B the rule was: "First render the divisor an integer by multiplying both dividend and divisor by 10 or some power of 10. Then proceed as with integral divisors." The conclusion favors Method B.

Winch⁴⁵ has reported that the "method of unity" is an effective method of teaching proportion. This method is illustrated in the following problem:

I pay 4 shillings for 2 pairs of boots. What shall I have to pay for 1 pair? What shall I have to pay for 3 pairs?

The use of the two questions in these problems directs the solution of the problem from the easy to the more difficult. Winch also reported that proportion in its simpler forms may be taught to children as young as seven years of age, that there do not appear to be any clear sex differences in ability to handle proportion, and that vacation seemed to have little effect on the proportion abilities. He states the very interesting conclusion: "The pupils of schools of very low social class—'slum schools'—cannot, even in the most favorable

⁴²Clapp, F. L., Chase, W. J., and Merriman, Curtis. "A Study of the Effectiveness of Two Kinds of Teaching Material," *Introduction to Education*. Boston: Ginn and Company, 1929, p. 420-24. (25)

⁴³Monroe, W. S. "The Ability to Place the Decimal Point in Division," *Elementary School Journal*, 18:287-93, December, 1917. (77)

⁴⁴Drushel, J. A. "A Study of the Amount of Arithmetic at the Command of High-School Graduates Who Have Had No Arithmetic in Their High-School Course," *Elementary School Journal*, 17:657-61, May, 1917. (35)

⁴⁵Winch, W. H. "Should Young Children Be Taught Arithmetical Proportion?" *Journal of Experimental Pedagogy*, 2:79-88, 319-30, 406-20; June, 1913; June 5, December 5, 1914; 3:89-95, June 5, 1915. (126)

pedagogical circumstances, be expected to undertake the work at as early an age as the others."

Springer⁴⁶ conducted an experiment in which the effectiveness of memorizing tables of cubic and linear measure was compared with the effectiveness of using the facts of these tables in connection with problems. The conclusions favor isolated memorizing of denominate-number facts rather than attempting to learn them in connection with the solving of problems in which they occur.

2. **Evaluation of the experiments.** The studies of Collier (30); Clapp, Chase, and Merriman (25); Winch (126); and Springer (109) were experimental in nature. Collier (30) used two groups of fifth-grade children, each of which numbered four individuals. No attempt was made to secure equivalence, and the experiment lasted only five days. It was observed that the experimental pupils learned to multiply fractions more quickly than did the control pupils. It is evident that this was a very crude experiment. Its faults are many: small groups, lack of equivalence, short duration, inadequate measurement of gains, and so on. The conclusion that children should be taught to multiply fractions through addition of fractions seems reasonable, but Collier's evidence in support of this conclusion is of doubtful value.

Clapp, Chase, and Merriman (25) employed twenty-three pairs of groups of unreported size. Both groups of a pair were taught in the same room by the same teacher. Equivalence was sought with respect to intelligence and initial ability in arithmetic. The duration of the experiment is not stated. At the end of the experiment three tests of eight percentage problems and two other problems each were administered. The results in twenty out of the twenty-three rooms favored the experimental factor—the novel percentage practice material. Clapp, Chase, and Merriman are to be commended for using so many pairs of groups, for attempting to secure equivalence with respect to two important pupil characteristics, and for the attempt to control non-experimental factors by having the same teacher instruct both experimental and control children. Instruction of a pair of classes by the same teacher, however, does not necessarily insure complete control of the non-experimental factors. Since the practice material was novel, it would not be unreasonable if there was some lack of equivalence in the teacher factors of zeal and effort. Furthermore, the merit of the experiment is possibly obscured by the method of reporting. One wishes for data relative to the sizes of the groups, to the degree of equivalence secured, and to the differences in gains in

⁴⁶Springer, Isidore. "Teaching Denominate Numbers," *Journal of Educational Psychology*, 6:630-32, December, 1915. (109)

achievement along with measures of the "statistical" significance of these differences.

The report by Winch (126) refers to five single-group experiments and one control-group experiment. The single groups varied in size from 39 to 361. The smallest group was located in a school in a good district; the rest were located in schools in the poorer districts of London, England. There was no attempt in the single-group experiments to control non-experimental factors. These experiments lasted from three to five months in the different groups. At the close of each experimental period, informal tests were administered and the improvement was noted. In the controlled experiment, two groups of twenty-three English school girls averaging nine years of age were equated with respect to initial arithmetical ability, as revealed by a series of preliminary tests. One group was taught in the usual fashion, while the other group was instructed in proportion by the method of unity. After three practice periods of 17, 16, and 22 minutes' duration, two of the preliminary tests were repeated. The difference in achievement favors the method of unity, but since this difference is but 2.5 times its probable error, it may not be regarded as "statistically" significant. Winch is to be commended for his careful analysis of the method of instruction used, for repeated experiments, and for his attempts to allow for the influences of non-experimental factors even where control groups were not used.

Springer (109) used two sixth-grade groups of fifty pupils each. Equivalence was secured with respect to initial ability in arithmetic as revealed by a test of arithmetical problems and with respect to language ability as shown by a language test. The experimental factor does not appear to have been adequately defined and isolated, and, although the groups were rotated, the control of the non-experimental factors was not satisfactory. The experiment is also to be criticized for its short duration—six periods of ten minutes each. The differences in achievement in favor of the isolated learning of the denominate-number facts appear to be fairly significant, although no standard or probable errors are given. The experiment is to be criticized for its failure to secure adequate control of non-experimental factors, as well as for its short duration.

Monroe (77) collected his data relative to the abilities required in placing the decimal point in division by means of four tests lasting one minute each, which were administered to seventy-eight sixth-, seventh-, and eighth-grade pupils. Anspaugh (2) merely reports what happened in a few elementary schools as a result of greater attention to the mastery of the fundamental number combinations.

His study may be termed "experimental" only in the sense any trial of a new method is experimental. Drushel (35) collected his data by the administration of his test to 624 entering college freshmen. The test results revealed that the method in which the divisor is rendered an integer by multiplying both dividend and divisor by 10 or by some power of 10 is significantly the better method. Considering the number of cases on which it is based, the "statistically" significant differences in achievement, and the approximate equivalence of the groups in general arithmetical ability, this conclusion seems quite dependable. This investigation, however, is not an experiment and consequently the degree of control of non-experimental factors is unknown. Hence, the superiority of "rendering the divisor an integer by 10 or some power of 10" cannot be said to have been demonstrated.

3. **Justified conclusions.** The crudity of the experiments described prevent the listing of justified conclusions.

CHAPTER III

DRILL IN THE FUNDAMENTALS

Consideration is given first in this chapter to the experiments which have been conducted for the purpose of revealing the effect of drill in the fundamentals. Attention is given next to the relative merits of systematic and incidental instruction in calculation. This is followed by a summary of the investigations in which the type of learning exercises was made the experimental factor. The chapter closes with an evaluation of the research on methods of distributing practice time in arithmetical calculation and on the influence of requests for speed and for accuracy on achievement in the fundamentals.

THE EFFECT OF SYSTEMATIC DRILL IN THE FUNDAMENTALS

1. Summary of reported conclusions. Studies of the effect of a period of systematic drill on achievement in arithmetical calculation¹ have produced evidence in support of the wide-spread belief that ability to add, subtract, multiply, and divide may be increased by systematic drill. Hagen² is the only investigator whose findings are not in entire agreement with this belief.

2. Evaluation of the experiments. Although Brown's study (13) is the earliest of this group, the technique used seems to have been superior to the techniques of any of the later experiments. In the first of Brown's studies, two groups of twenty-five sixth-, seventh-, and eighth-grade children were paired on the basis of their initial ability in arithmetic. The arithmetic instruction of one of the groups differed from that of the other in that five minutes of each of thirty recitation periods were devoted to drill in the four fundamentals. At the end of the experiment, a final test, similar to the initial test by which the groups were equated, was administered. The second exper-

¹Brown, J. C. "An Investigation on the Value of Drill Work in the Fundamental Operations of Arithmetic," *Journal of Educational Psychology*, 2:81-88, February, 1911; 3:485-92, 561-70; November, December, 1912. (13)

Burton, C. B. "Results of Definite Drill in the Four Fundamental Processes as Shown by the Woody-McCall Mixed Fundamentals," *Fifth Yearbook of the Department of Elementary School Principals*. Washington: National Education Association, 1926, p. 323-28. (19)

Kerr, M. A. "Effects of Six Weeks Daily Drill in Arithmetic," *Studies in Arithmetic, Indiana University Studies* No. 32. Bloomington: Indiana University, 1916, p. 79-95. (56)

Phillips, F. M. "Value of Daily Drill in Arithmetic," *Journal of Educational Psychology*, 4:159-63, March, 1913. (100)

Smith, J. H. "Individual Variations in Arithmetic," *Elementary School Journal*, 17:195-200, November, 1916. (107)

Wimmer, H. "Experimental Study of the Effects of Drill in Arithmetic Processes under Varying Conditions," *Indiana University Studies* No. 32. Bloomington: Indiana University, 1916, p. 96-102. (124)

²Hagen, H. H. "A Study of Practice Periods in Arithmetic Fundamentals," *Chicago Principals' Club, Second Yearbook*. Chicago: Chicago Principals' Club, 1927, p. 93-95. (45)

iment was similar to the first with the exception that 222 pupils in four schools participated for twenty recitation periods. Brown is to be commended for the techniques which he used in securing equivalent groups, for his care in controlling non-experimental factors, and for his elaborate analysis of the data. He is also to be commended for repeating his experiment with pupils in several schools and in different cities. His differences in gains in achievement, secured in this way, are of sufficient magnitude to support adequately his conclusion with respect to the effect of systematic drill of five minutes per day on achievement in arithmetical calculation.

The other five studies of the effect of systematic drill are subject to criticism. Kerr (56) used 423 sixth-, seventh-, and eighth-grade children in her single-group experiment. These children received five minutes of drill in addition, daily, for a period of six weeks. The application of an initial and a final test showed a gain in ability to add, but the significance of this gain is obscured by the failure of the experimenter to employ a control group. Phillips (100) used two groups of thirty-four and thirty-five sixth-, seventh-, and eighth-grade children. After these pupils had been paired on the basis of initial ability in arithmetic, the members of the experimental group were given ten minutes of daily drill in the fundamental operations and with reasoning problems (mental arithmetic). At the end of two months the final test showed a "statistically" significant gain for the drill group. The techniques employed by Phillips seem much superior to those employed by Kerr (56), but his experiment does not seem to be without fault. The size of his groups was small, and the instructional conditions were not entirely normal. Smith (107) used three fifth- and sixth-grade classes of unreported size. No attempt was made to secure equivalence. One class received what amounted to diagnosis and remedial treatment during drill. The second class received extra drill for the inferior pupils. The third class was merely drilled. After three drill periods per week of twenty-five minutes each for four weeks the final tests were administered. The magnitude of the gains in achievement seems to warrant the statement: "All three types of drill produced very large increases in the achievement of the pupils." The conclusions which state that the first type of drill is significantly superior to the other two would seem to be less dependable. Smith is to be criticized for failure to secure equivalent groups, for evidently poor control of the time factor, and for failure to report the size of his groups. With respect to the comparative value of drill, this must be regarded as a single-group or uncontrolled experiment.

Wimmer (124) employed fifth-, sixth-, seventh-, and eighth-grade pupils. The pupils in the sixth grade were divided into two apparently equivalent groups of twenty-two pupils each. The other classes which averaged about thirty-five pupils each were used as single groups. The Courtis Standard Test, Series A, was administered at the beginning of the experiment, at the end of six weeks, and again at the close of the experiment—twelve weeks from the beginning. Comparisons are made between the gains of the different classes and between the two groups of the sixth-grade class. The classes which had systematic drill made the greater gains, but the magnitude of the differences in gains is obscured by faulty or complete lack of equivalence. The gains are large enough, however, for the classes which had drill to justify the conclusion that "it pays to give regular drill work in arithmetic."

Burton (19) employed 2500 third-, fourth-, fifth-, sixth-, seventh-, and eighth-grade pupils in the white rural schools of a county in one southern state. Systematic drill was administered ten minutes daily for a period of six weeks. Curves are given to show the consistent gains in efficiency made by the pupils. The experimenter is to be commended for the large number of pupils used, but he is to be criticized for not using some of the pupils for control purposes.

Hagen (45) employed twelve pairs of groups of fourth-, fifth-, sixth-, and seventh-grade pupils which were equated on the basis of intelligence test scores. Each teacher participating in the experiment taught a pair of groups. One of the groups of each pair received systematic drill in fundamental problems twice each day, while the other group received the drill once a day. After three months of such instruction the final test was administered. The difference in achievement, when the gains of all the groups are averaged, slightly favors drill once a day. That this difference is not of much significance is shown by the fact that in six of the twelve pairs of groups the mean differences in achievement slightly favor the use of drill twice a day. The following statement of Buswell relative to the experiment seems justified: "Data might be interpreted differently."³

3. **Justified conclusions.** If the Law of Exercise is accepted, it is obvious that pupils who have not attained their maximum skill in arithmetical calculation will profit from systematic drill, especially when the drill is conducted in a way that stimulates a desire to increase achievement in this field. Consequently this group of six studies may be labelled as "attempts to prove the obvious." The

³Buswell, G. T. "Summary of Arithmetic Investigations," *Elementary School Journal*, 28:705, May, 1928.

conclusions, except possibly certain incidental details, are merely what should have been anticipated.

THE RELATIVE VALUE OF SYSTEMATIC VERSUS INCIDENTAL TEACHING OF CALCULATION

1. **Summary of reported conclusions.** Meriam⁴ and Collings⁵ secured results that favored incidental teaching of calculation, but Gates, Batchelder, and Betzner⁶ have reported that the differences in arithmetic achievement in their experiment favored the "systematic" rather than the "opportunistic" method of instruction. Wilson⁷ has reported recently that incidental instruction of the informational type is just as effective as instruction of the traditional type, so far as the first two grades are concerned, and that a combination of both types with more emphasis on systematic drill results in very superior arithmetical achievement in the third grade.

One of the conclusions of the investigation recently reported by Olander (94) may be interpreted in favor of systematic teaching of calculation:

Examination of the scores of one group of children who had no formal instruction in arithmetic for twelve out of the seventeen weeks of the experiment and of another group who had no formal arithmetic instruction whatsoever during the entire seventeen weeks shows that, during the time when no class instruction in numbers was being given, the children learned from approximately a third to less than a half as many number combinations as did the children who were being given the regular class instruction.

2. **Evaluation of the experiments.** Meriam (73) merely reported a comparison of grades in high school of 362 pupils who had received incidental instruction in arithmetic, in the elementary school, with the grades of those who had had the more traditional form of instruction. The findings of such an investigation cannot be accepted as conclusive, in any sense. There were too many factors unaccounted for which may have influenced the results.

Collings (31) used forty-one pupils in one rural school as his experimental group and sixty pupils in two other rural schools as his control group. The initial arithmetic test revealed the fact that the experimental pupils were slightly inferior to the control pupils in ability in the four fundamentals. Collings also presents much evidence relative to the approximate equivalence with respect to reading ability, hand-

⁴Meriam, J. L. "How Well May Pupils Be Prepared for High School Work without Studying Arithmetic, Grammar, etc., in the Grades?" *Journal of Educational Psychology*, 6:361-64, June, 1915. (73)

⁵Collings, Ellsworth. *An Experiment with a Project Curriculum*. New York: Macmillan Company, 1923. 346 p. (31)

⁶Gates, A. I., Batchelder, M. I., and Betzner, Jean. "A Modern Systematic versus an Opportunistic Method of Teaching," *Teachers College Record*, 27:679-700, April, 1926. (40)

⁷Wilson, G. M. "New Standards in Arithmetic: A Controlled Experiment in Supervision," *Journal of Educational Research*, 22:351-60, December, 1930. (123)

writing ability, spelling ability, chronological age, number of years of schooling, number of years spent in the experimental schools, school attitudes, community attitudes, social and economic status of the districts, parentage of children, length of school term, course of study, and so on. After four years of the project curriculum in the experimental school and four years of the traditional curriculum in the control schools the final tests were administered. With respect to ability in the four fundamentals, the differences favor, but not significantly, the informal method. Collings has been criticized for his failure to control important non-experimental factors:

In the experiment by Collings the children taught by the project method achieved more than those taught by the traditional method, but it appears from Collings' report that these teachers worked much harder at their task than did the teachers in the control schools. In view of this fact, it does not appear justifiable to ascribe the superior achievement of the project-method group entirely to the method of instruction.⁸

Gates, Batchelder, and Betzner (40) employed two groups of twenty-five first-grade children who were approximately equivalent with respect to such traits as sex, chronological age, mental age, general information, speed of reading, oral spelling, and so on. The group subjected to the opportunistic method was somewhat inferior to the other group in initial ability in oral arithmetic. Techniques used to control teacher factors are described in the following quotation:

Both teachers were interested in the project as an experimental study; both, understanding that the results would in no way reflect upon their professional reputation, taught their pupils as under ordinary circumstances except for certain imposed limitations and regulations which were cheerfully accepted and faithfully observed. Both teachers followed the same general schedule, the same time assignment to different phases of the work, recesses, lunch periods, assembly music, gymnasium work, and so forth. Neither teacher gave any out-of-school time to individual pupils nor allowed others to do so; neither suggested home work, and each as far as possible, prevented it. Neither was given any assistance in teaching; neither enjoyed any advantage in clerical or other help, in funds for materials, in special demonstrations, and so on.

It is the opinion of the present writers that the techniques used to control the teacher factors and the other non-experimental factors in this experiment were superior to those used by Collings (31). "Each of the two methods, 'the modern systematic' and the 'opportunistic,' was followed by an exceptionally able teacher who was experienced in the method and believed it to be, on the whole, the best one."⁹

⁸Monroe, W. S. and Engelhart, M. D. "Experimental Research in Education," *University of Illinois Bulletin*, Vol. 27, No. 32, Bureau of Educational Research Bulletin No. 48. Urbana: University of Illinois, 1930, p. 36.

⁹Gates, Batchelder, and Betzner, *op. cit.*, p. 682.

If this was the case, it would seem that the teacher factors, skill and zeal, were rather adequately controlled.

The difference in achievement, as revealed by the final test in arithmetic at the end of the year, was 2.5 times the probable error of the difference. As such, the difference may be regarded as possibly "statistically" significant. A limitation of this experiment, so far as arithmetic is concerned, is the lack of equivalence in arithmetic ability at the beginning of the experiment. Some of the difference in the final achievement in arithmetic may be attributed to the initial superiority of the systematic group. Hence, it appears that the difference should not be interpreted as more than suggestive.

Wilson (123) compared the scores of 475 pupils completing the second grade, who had received informal or incidental instruction in arithmetical calculation, with the scores of one group of 174 second-grade pupils and one group of 154 third-grade pupils, who had received the traditional formal type of instruction. These data support the contention that up to the close of the second grade the informal type of arithmetical instruction results in achievement equal to, and possibly superior to, the achievement resulting from formal instruction. In the later phases of Wilson's experiment over one thousand third-grade children were subjected to a combination of incidental and systematic instruction. One day a week during the third year was devoted to incidental instruction of the informational type, while the other four days were devoted to systematic drill on addition and subtraction. The tables of test results indicate that the pupils attained a very high level of achievement in addition and subtraction. While Wilson's conclusion seems well supported by his data, one wonders whether too much emphasis was not placed on the informal aspect of the instruction and too little recognition given to the part played by systematic drill in securing the superior achievement of the third-grade children.

In Olander's experiment (94) one group of one hundred second-grade pupils received no instruction in arithmetic for the last twelve of the seventeen weeks of the experiment.¹⁰ Another group of eighty-six pupils received no formal arithmetic instruction during the entire seventeen weeks. The achievements of these groups were compared with each other and with the achievement of a group of 296 pupils receiving daily instruction. The initial ability of the group of eighty-six was considerably superior, and that of the group of one hundred, slightly superior, to the initial ability of the group receiving daily

¹⁰See pages 17 to 18 for evaluation of this experiment, which had to do with the effectiveness of generalization instruction.

instruction. It would seem, therefore, that the differences in favor of systematic daily instruction are rather highly reliable.

3. **Justified conclusions.** The conflicting conclusions of the experiments evaluated prevent the formulation of a justified conclusion favoring either the incidental or the traditional method of instruction in arithmetic. The question as to which method is superior awaits further experimental investigation. In view of the relatively specific character of calculation abilities and the demonstrated efficacy of systematic drill, it is difficult to conceive of the incidental method alone as highly efficient. It is possible that the best method would be a combination of the two procedures.

THE RELATIVE MERITS OF CERTAIN GENERAL TYPES OF LEARNING EXERCISES FOR DRILL IN CALCULATION

1. **Summary of reported conclusions.** Ten experimental investigations are summarized under this heading. Evans and Knoche¹¹ have reported that drill in which Studebaker Economy Practice Exercises are used results in achievement superior to that resulting from the use of learning exercises based on materials devised by the teacher. Kelly¹² compared the effectiveness of the Curtis Standard Practice Tests, the Studebaker Economy Practice Exercises, and "the best methods of drill which the teachers could devise." He reported that the Curtis drill material is superior to the Studebaker material, but that both are superior to drills devised by the teachers. Mead and Johnson¹³ compared the Curtis Standard Practice Tests with the Thompson Minimum Essentials and reported a conclusion favorable to the Curtis material. Morgan¹⁴ compared the effectiveness of the Economy Remedial Exercise Cards when used with the Compass Diagnostic Tests to that of Lennes' Pads and reported a conclusion favorable to the former. Newcomb¹⁵ found that drill exercises prepared in such a way that proportionate drill is given on the higher decades are more effective than those ordinarily used. Fowlkes¹⁶ concluded that it is desirable "to teach the one hundred combinations (multiplication) by means of text material alone, the teacher doing as little talking as possible" and "to make remedial adjustments by means of printed directions and devices rather than

¹¹Evans, J. E. and Knoche, F. E. "The Effects of Special Drill in Arithmetic as Measured by the Woody and the Curtis Arithmetic Tests," *Journal of Educational Psychology*, 10:263-76, May-June, 1919. (37)

¹²Kelly, F. J. "The Results of Three Types of Drill on the Fundamentals of Arithmetic," *Journal of Educational Research*, 2:693-700, November, 1920. (55)

¹³Mead, C. D. and Johnson, C. W. "Testing Practice Material in the Fundamentals," *Journal of Educational Psychology*, 9:287-97, May, 1918. (71)

¹⁴Morgan, L. D. "Specific vs. General Drill in the Fundamentals of Arithmetic," *School Science and Mathematics*, 29:528-29, May, 1929. (80)

¹⁵Newcomb, R. S. "Effective Drill Exercises in Arithmetic," *Journal of Educational Psychology*, 16:127-31, February, 1925. (88)

¹⁶Fowlkes, J. G. "A Report of a Controlled Study of the Learning of Multiplication by Third-Grade Children," *Journal of Educational Research*, 15:181-89, March, 1927. (38)

oral instruction." Knight¹⁷ has reported a conclusion which favors "drill material carefully constructed as to the distribution of practice in addition, subtraction, multiplication, and division of whole numbers," rather than drill material "slightly in excess as to sheer amount but so built that certain combinations were slighted."¹⁸ The conclusions of Newcomb (88), Fowlkes (38), and Knight (60) all favor the contention that the relative difficulty of the number combinations must be accounted for in preparing efficient materials of instruction for use in drill.

Kulp¹⁹ investigated the relative effectiveness of two types of practice material, the essential difference between the two being that one of the types provided practice in solving reasoning problems in connection with computational drill. It is reported that the material which provided practice in arithmetical reasoning was relatively more effective in securing computational achievement, and that its use resulted in a decided increase in arithmetical reasoning ability. A similar conclusion is reported by Rosse.²⁰ These conclusions seem to agree with that reported by Kirkpatrick²¹ several years ago. Kirkpatrick found that use in calculation is a more effective means of learning the multiplication combinations than memorization divorced from use.

Myers and Myers²² investigated the problem of whether it was better to find mistakes among a group of examples of addition, multiplication, and subtraction combinations than to think of the corresponding correct associations. Their results are favorable to learning exercises which emphasize correct associations rather than learning exercises which demand the observation of errors. It is interesting to note, among their conclusions, that pupils thought the discovery of errors made by other people much more interesting than the drill in which correct associations were exercised.

The problem of whether learning exercises should be restricted to one arithmetical operation or should deal with more than one has been studied in three experiments. Buckingham²³ sought to determine whether it is "better to teach subtraction facts in connection

¹⁷Knight, F. B. "The Superiority of Distributed Practice in Drill in Arithmetic," *Journal of Educational Research*, 15:157-65, March, 1927. (60)

¹⁸Knight summarizes in this article the report of an experiment conducted by Luse. See:

Luse, E. M. "Transfer within Narrow Mental Functions, A Study of the Effects of Distributed versus Non-Distributed Drill in Arithmetic," *University of Iowa Monograph in Education* No. 5. Iowa City: University of Iowa. (61)

¹⁹Kulp, C. L. "A Study of the Relative Effectiveness of Two Types of Standard Arithmetic Practice Materials," *Journal of Educational Research*, 22:381-87, December, 1930. (65)

²⁰Rosse, J. C. "An Experiment to Test the Increase in Reasoning Ability from the Use of Test and Practice Sheets in 6A Arithmetic," *Journal of Educational Research*, 22:210-13, October, 1930. (105)

²¹Kirkpatrick, E. A. "An Experiment in Memorizing versus Incidental Learning," *Journal of Educational Psychology*, 5:405-12, September, 1914. (58)

²²Myers, G. C. and Myers, C. E. "Finding Mistakes versus Correct Associations in Simple Number-Learning," *Journal of Educational Research*, 18:25-31, June, 1928. (86)

²³Buckingham, B. R. "Teaching Addition and Subtraction Facts Together or Separately," *Educational Research Bulletin* (Ohio State University), 6:228-29, 240-42; May 25, 1927. (17)

with related addition facts than to teach the addition facts first and the subtraction facts afterward." His conclusions favor the teaching of addition and subtraction together. Myers and Myers²⁴ prepared learning exercises which required the pupils to shift rapidly among the four fundamental operations. Their conclusions are distinctly unfavorable to such mixed exercises. ". . . rapid shifting by the pupil from one process to another not only causes great confusion of processes, but the pupil so confused also tends to be more confused when he later works on combinations grouped twenty-five to a process." Repp²⁵ prepared two sets of drill material the objective of which was the maintenance of skill. Each of the exercises of one set of material dealt with a single topic, such as addition of fractions, while each exercise of the other set of material was of mixed nature. This difference in organization was the only difference in the content of the two sets of drill material. The conclusions are distinctly favorable to the mixed type of drill material as a basis of learning exercises for the maintenance of skills in arithmetic. "All pupils profited by use of drills furnished them, but those using mixed drills showed 23 per cent greater gain than those using isolated drills."

2. **Evaluation of experiments.** Evans and Knoche (37) used two groups of sixth-grade children of unreported size. With respect to equivalence they state that "the children in the two rooms were quite similar in ability. The 6A class was one semester in advance of the 6B group." The pupils of the 6B class were drilled with the Studebaker Economy Practice Exercises five minutes each day for forty-three days, the time being taken from their regular arithmetic work. The tests administered at the end of the experimental period yielded a probably significant difference in mean gain for the group using the Studebaker Exercises. The experimenters are to be criticized for not attempting to secure equivalent groups and for utilizing pupils whose arithmetic instruction, other than that inherent in the experimental factor, differed so greatly. It is stated that during the period of drill "the main work for the 6A grade was percentage with a general review of the fundamental processes. The work of the 6B grade was decimals." It is possible that the zeal of the teacher for the novel practice material was another uncontrolled factor.

Kelly (55) used three groups of 133, 146, and 173 fourth-, fifth-, sixth-, seventh-, and eighth-grade children, making no effort to secure equivalence. The groups used the Courtis Standard Practice Tests,

²⁴Myers, G. C. and Myers, C. E. "The Cost of Quick Shifting in Number Learning," *Educational Research Bulletin* (Ohio State University), 7:327-34, October 31, 1928. (85)

²⁵Repp, A. C. "Mixed versus Isolated Drill Organization," *Twenty-Ninth Yearbook of the National Society for the Study of Education*, Bloomington, Illinois: Public School Publishing Company, 1930, p. 535-49. (103)

the Studebaker Economy Practice Exercises, or informal exercises prepared by the teachers for eight to fifteen minutes of drill per day, depending on the grade level, for twenty successive days. The techniques used in this experiment are open to criticism. A lack of equivalence is indicated by the unequal representation of the different school grades in each of the groups. For example, there were no fourth-grade children in the group using the Curtis material and no VA or VIB pupils in the group using the Studebaker material. Failure to control important teacher factors is indicated in the statement that "The differences from class to class by the same method suggest that after all the efficiency of any method depends mostly on the teacher who is using it."

Mead and Johnson (71) used two groups of 105 fifth- and sixth-grade pupils. No attempt was made to secure equivalence, and the preliminary tests reveal some departures from equivalence. The pupils of one group practiced ten minutes a day with the Curtis material, while the pupils of the other group used the Thompson material. No attempt was made to prevent home practice, it being felt by the experimenters that if a practice material stimulated such practice such stimulation should be allowed to operate during the experiment. After ninety days of practice the Curtis Research Test was administered, the results of which were possibly significantly in favor of the Curtis Standard Practice Tests. This experiment is faulty in that no effort was made to secure equivalence or to control practice time. Precision in experimentation demands that pupils of experimental and control groups spend an equal amount of time in learning. Another possible fault is that the Curtis Research Test would be more valid with respect to the Curtis drill material than with respect to the Thompson drill material.

Morgan (80) used two groups of twenty-eight fourth-grade pupils. The groups were equated on the basis of average scores made on two standardized arithmetic tests. One group used the Economy Remedial Exercise Cards and was subjected to the Compass Diagnostic Tests, while the other group merely used practice pads prepared by Lennes. Both groups were taught by the same teacher for a period of twelve weeks. At the end of this period, the other forms of the initial tests were administered, and the average scores, computed. The difference in mean gains significantly favors the group that used the Economy Remedial Exercise Cards and that had the Compass Diagnostic Tests administered to it. There seems little reason to doubt the reliability of the findings, but it is impossible to ascribe the superior achievement of the group which excelled to the practice material

or to the diagnostic tests. It would seem, therefore, that the chief criticism which may be made with respect to this experiment has to do with the failure of the experimenter to restrict the experimental factor to a single technique.

Newcomb (88) used an experimental group of fifty-one pupils and a control group of twenty-one seventh-grade pupils. With respect to equivalence he states, "A comparison of the intelligence quotients of the pupils of the several classes did not reveal on the whole any appreciable differences." The experimental group was practiced five or six minutes a day for thirty-five days on drill material which provided practice on the higher decades, while the instruction of the control group was conducted "in the usual manner." The administration of the Courtis Standard Research Test at the close of the experiment revealed the probably significantly superior achievement of the experimental group. Newcomb is to be criticized for not securing more adequate equivalence of groups and for not specifying the type of learning activity engaged in by the control pupils. It is possible that greater zeal was exerted by the teachers in utilizing the experimental drill material, since the failure to mention the type of drill material used by the control pupils would indicate a lack of enthusiasm for it.

Fowlkes (38) used a single group of thirty-one third-grade pupils whose median I. Q. was 104.5. This group of pupils was drilled on multiplication twenty minutes a day for twenty days by means of the text material alone, "the teacher doing as little talking as possible," and remedial adjustments were made by printed directions and devices. There resulted from this instruction achievement which is claimed by the author to be significantly better than that of other third-grade classes. While a single-group technique is not usually to be relied upon, the fact that Fowlkes was able to compare his results with those of other third-grade classes would give his conclusions some dependability. It is possible that he should have allowed for the somewhat superior intelligence of his third-grade class in formulating his conclusions.

Luse, as reported by Knight (60), used two groups of three hundred fifth-grade pupils which were equivalent with respect to general arithmetic ability. One of these groups used carefully constructed material, while the other employed material which slighted certain of the number combinations. "All other conditions were held constant." After fifty consecutive drill periods of fifteen minutes each, the final tests were administered. The differences in achievement were probably "statistically" significant in favor of drill material in which practice is carefully distributed over the number combinations. The

techniques used in this experiment compare favorably with the best of contemporary experimental research in education.

In the experiment of Kulp (65) four classes used the practice material which did not provide practice in arithmetical reasoning, while six classes used the material which did. A total of 113 fourth-grade pupils took the final test. It is evident from the figures given in the report of the investigation that the experimental and control groups were initially equivalent in computational ability, but that the group receiving the training in reasoning was initially superior in reasoning ability. The teacher factor, experience with instructional procedure, favored the practice material which did not provide practice in solving reasoning problems, but it is possible that the influence of this experience was offset by the usually occurring greater zeal for a new method or procedure. The experiment lasted from October to April. The differences in gains in achievement are apparently significantly in favor of the type of material which provided practice in arithmetical reasoning in connection with calculation drill. The investigator is to be criticized for failure to secure more adequate equivalence at the beginning of the experiment, and for failure to indicate more clearly the differences in gains in achievement and the "statistical" significance of these differences. The investigator is to be commended for his careful description of the compared factors, for measures taken to control non-experimental factors, and for conducting his experiment over a comparatively long period of time. His conclusions would seem to be fairly dependable with respect to the groups used in the experiment. Further experimentation is needed before generalization is justified.

Rosse (105) used two groups of eighteen sixth-grade pupils which were equivalent with respect to initial arithmetic reasoning ability and with respect to intelligence as measured by the Otis Arithmetic Reasoning Test and the National Intelligence Test. One group used practice sheets which provided drill in reasoning problems, while the other group used an ordinary arithmetic text. At the end of fifty-eight days the same form of the Otis Arithmetic Reasoning Test was administered. The difference in achievement favors, but not significantly, the method in which the practice sheets which provided drill in reasoning problems were used. While the conclusions do not seem to be highly dependable because of the size of the groups used, because of the lack of control of important non-experimental factors, and because of the unreliability of the difference reported, they may be accepted as evidence supplementing that presented by Kulp (65).

Kirkpatrick (58) used two groups of ten and two groups of twenty-

five normal-school students and two groups of twenty sixth-grade pupils, making no attempt to secure equivalence of groups. No mention is made of any procedures used to secure control of non-experimental factors. The groups were tested at the end of ten days, and the normal-school students, again at the end of three weeks. The differences in achievement in each case favored the method of learning multiplication combinations through use. It is evident that this experiment may not be regarded as other than crude. Since no attempt was made to secure equivalence of groups, or to control adequately non-experimental factors, the differences in achievement may not with certainty be ascribed to the method reported superior.

Myers and Myers (86) used two groups of one hundred fourth- and fifth-grade pupils which were matched on the basis of initial arithmetic ability. These groups were also matched with other groups of equal size in order to control the practice effect of the initial test. The experiment was conducted just long enough for the pupils of one group to observe errors in the answers of a group of twenty number combinations, while the members of the other group examined twenty combinations and their correct answers. The difference in achievement, as shown by the final test, was probably significantly in favor of the exercise in which the pupils observed only correct answers. The chief criticism of this experiment is its short duration. It is possible that the confusion caused by the exercise containing errors might have worn off with more prolonged use and that, in the long run, its use would result in superior achievement. It may be true, also, that this type of exercise is one which would engender the ability to locate mistakes—a well recognized objective of arithmetic instruction.

Buckingham (17) equated seven pairs of groups of from twelve to twenty-eight second-grade children in seven schools on the basis of scores on the Pressey Primary Classification Test. During a daily period of twenty minutes one of the groups of a pair was taught related addition and subtraction facts together, as for example: $1 + 6$, $6 + 1$, $7 - 1$, and $7 - 6$. The other group of pupils was taught all of the addition facts and then all of the subtraction facts for the same time per day. With the exception of this difference in the learning exercises, the instructional materials and techniques used for each pair of groups were the same. No home work was required and no new topics were introduced in arithmetic during the experimental period. The hour of the instruction was alternated for each pair of groups at the end of each week. The statement is made that the experiment lasted about a month for one of the pairs of groups,

but nothing is said in this respect about the others. Three of the differences in achievement revealed by the final test are "statistically" significantly in favor of the "together" method, and three more of the differences are in favor of the together method, but not significantly so. One difference favors, but not significantly, the separate method. Buckingham attaches great significance to this "all but unanimous verdict." He states, "When an experiment conducted seven times yields six results all in the same direction, the evidence is rather conclusive even though some of the differences, when considered individually, are small or lacking in statistical significance." He recognizes the limitation of the short duration of his experiment and the failure to test retention. While the techniques used in this experiment have some admirable features, a question may be raised with respect to the validity of the final test. Was it adapted to the type of learning exercises used by the different groups? If its examples were of mixed nature, it is probable that the test was more valid with respect to the mixed learning exercises. If, however, one of the groups, of a pair, had a test in which addition and subtraction were kept separate while the other group had the same items mixed, the results would probably be more valid with respect to each group, but it is difficult to see how they could be considered comparable. In the face of this dilemma of measurement one does not seem justified in accepting the conclusions as highly dependable.

Myers and Myers (85) used fifty fifth-grade pupils, sixty-four sixth-grade pupils, and fifty normal-school girls selected in a random fashion. "The first pupil of a given group was tested with the grouped combinations followed by the mixed combinations; the next pupil was tested with the mixed examples first and then with the grouped examples; the third pupil began with the grouped examples and so on alternating throughout the group." The pupils made their responses orally, and the experimenter recorded the time required. An analysis was made of the results, and a check was made of the practice effect. The results significantly favor the method of grouped, rather than the method of mixed, exercises.

Applying the two types of exercises to alternate pupils does not insure that they were applied to equivalent groups.²⁶ Another criticism concerns the length of the tests, each of which contained forty items. More dependable results could have been secured by the utilization of a much longer test, or by the utilization of a long period

²⁶This technique is probably justified when the groups are very large. For example, Monroe used a similar technique, but with a total of 9,256 pupils. See:

Monroe, W. S. "How Pupils Solve Problems in Arithmetic," *University of Illinois Bulletin*, Vol. 26, No. 23, Bureau of Educational Research Bulletin No. 44. Urbana: University of Illinois, 1929. 31 p. (79).

of learning prior to a final test. However, if this were done, the experimenter would yet be faced with the dilemma of a choice between a doubtfully valid mixed test or non-comparable separate tests.

Repp (103) used groups of 263 and 267 twelve-year-old pupils which were equivalent with respect to arithmetical ability as shown by an initial test of $.97 \pm .006$ reliability. One of these groups used drill material consisting of twenty-six twenty-minute exercises, each of which dealt with one topic. The other group used material of the same total content but of mixed organization. After twenty-six weeks an exhaustive final test, also of $.97 \pm .006$ reliability, and of a mixed nature was administered. The results of this test are "statistically" in favor of the mixed drills. The final test probably was more valid with respect to the abilities engendered by the mixed drills than with respect to the abilities engendered by the isolated drills. It should be mentioned, however, that an analysis of the achievement during practice ultimately favored the mixed drills. The conclusion may be justified, therefore, that mixed drills are superior for maintenance of skill, while isolated drills are superior in the earlier stages of learning.

3. Justified conclusions. If one accepts the principle that arithmetical ability in the field of calculation is specific, or at least largely so, and that, consequently, ability to calculate consists of a large number of specific abilities, it follows that drill must be provided on each specific ability, unless it is believed that there is essentially complete transfer from one specific ability to another when these abilities are at all closely related.²⁷ Furthermore, it appears reasonable that the more difficult combinations should receive more drill than the easier ones. Consequently, it is to be expected that learning exercises constructed with due recognition of the specific abilities to be engendered and of their relative difficulties and interrelations should be more effective than learning exercises not so constructed. This group of investigations supports this general hypothesis and appears to justify the assertion that the hypothesis has been demonstrated. It might be argued that this hypothesis is obvious and, hence, that the principal contribution of these studies is to be found in their details. The more significant of these detailed findings appear to be:

²⁷The conclusions of the recent investigations of Beito and Brueckner (9) and of Olander (94) would seem to indicate that there is a large amount of transfer in the case of certain abilities. The conclusions of Beito and Brueckner (9) were referred to in a footnote on page 16. Olander (94) has reported that "The ability gained by children on fifty-five simple number combinations in addition and on fifty-five similar combinations in subtraction transferred almost completely to the forty-five remaining simple number combinations in each of the two processes." This conclusion seems to be reasonably dependable, since Olander used relatively large equivalent groups, controlled non-experimental factors rather adequately, and secured measures of achievement which seem acceptably reliable and valid. Such a conclusion would not seem to oppose the contention above that the best materials for drill are those constructed so that the more difficult combinations receive the greater practice. It is commonly accepted as a principle in education that the best way to insure attainment is to practice the needed abilities directly rather than to depend on transfer.

1. Practice material prepared by experts seems to be more effective than learning exercises based on material prepared by teachers.
2. Learning exercises in which the practice is carefully distributed over the number combinations so that none are slighted and the more difficult combinations occur with relatively greater frequency are superior to learning exercises which have not been thus prepared.
3. Learning exercises to be used in the initial stages of learning calculation should probably require the practice of addition, subtraction, multiplication, and division separately. Learning exercises whose objective is the maintenance of skill should be mixed in character. The pupils should be given some opportunity to practice their calculation abilities in the situations represented by examples varied with respect to the fundamental process called for.

THE INFLUENCE OF DISTRIBUTION OF PRACTICE TIME ON ACHIEVEMENT IN THE FUNDAMENTALS

1. **Summary of reported conclusions.** Three experiments have been reported on the effect of distribution of practice time on learning, and one has been reported on the distribution of practice needed for retention, or maintenance of skill. Kirby²⁸ compared practice periods in addition of 22½, 15, 6, and 2 minutes' duration and in division of 20, 10, and 2 minutes' duration. The gains in achievement, for both addition and division, favored the two-minute interval. Hahn and Thorndike²⁹ compared practice periods in addition of 5, 7½, 10, 11¼, 15, 20, and 22 minutes' duration. Their results tend to favor the longer periods. Wimmer (124) reported that pupils who were given one fifteen-minute drill per week made greater progress than those who were given five minutes of drill five times per week. Reed³⁰ compared a single hour of practice in addition with a distribution of twenty minutes a day for three days, ten minutes a day for six days, and ten minutes twice a week for three weeks. The gains in achievement favor the distribution of twenty minutes a day for three days.

Norem and Knight³¹ investigated the distribution of practice needed for retention or maintenance, of skill. They concluded with respect to drill in multiplication that when mastery has been attained "one practice a week is sufficient for maintenance." They state also, however, that one practice a week "is often insufficient practice for maintaining the combinations during the first two weeks following the initial learning of them."

²⁸Kirby, T. J. "Practice in the Case of School Children," *Teachers College, Columbia University Contributions to Education*, No. 58. New York: Bureau of Publications, Teachers College, Columbia University, 1913. 98 p. (57)

²⁹Hahn, H. H. and Thorndike, E. L. "Some Results of Practice in Addition under School Conditions," *Journal of Educational Psychology*, 5:65-84, February, 1914. (46)

³⁰Reed, H. B. "Distributed Practice in Addition," *Journal of Educational Psychology*, 15:248-49, April, 1924. (102)

³¹Norem, G. B. and Knight, F. B. "The Learning of the One Hundred Multiplication Combinations," *Twenty-Ninth Yearbook of the National Society for the Study of Education*, Bloomington, Illinois: Public School Publishing Company, 1930, p. 551-68. (91)

2. **Evaluation of the experiments.** Kirby (57) employed groups of 194, 104, 205, and 229 fourth-grade children in his addition experiment. These groups were practiced fifteen minutes in addition as an initial test. They were then subjected to forty-five minutes of practice divided into periods of $22\frac{1}{2}$, 15, 6, or 2 minutes in length. Finally, they were practiced for another fifteen-minute interval, which represented the final test. The experimenter exercised considerable care to prevent the children from practicing outside of the practice intervals and to control other non-experimental factors. He conducted the practice himself in practically all of the classes. The experiment with practice divided into periods of 20, 10, and 2 minutes' duration in division was conducted in a similar fashion, using groups of 204, 209, and 193 third- and fourth-grade children. The differences in gains seem possibly significant with respect to addition practice periods of two minutes' duration and certainly significant with respect to division practice periods of the same length.

Kirby is to be commended for his attempt to secure a representative sample of school children. He checked the performance of thirty-eight of the school classes which were used in this experiment, and which were located in New York City, with results obtained with a class outside of this city. One fault to be found with this experiment is that of failure to secure equivalent groups. While the failure to secure equivalence does not invalidate the results, it does obscure their precise significance. The experimenter calls attention to the possible influences of factors not inherent in the short practice period:

(1) The groups, working in shorter periods, because of the number of days over which the experiments ran, had greater opportunity during the experiment to profit from the regular school work than other classes (2) The groups working in shorter periods had a longer time in which to catch the spirit of the experiment and to become enthusiastic over surpassing their previous performance. They had their records read to them more times and had the incentives to intense effort repeated more often. (3) They also had greater opportunity and incentive to do work outside of the time given to the experiment.

The experiment of Wimmer (124) was described and evaluated on page 32. His conclusion with respect to the distribution of practice time may not be regarded as dependable.

Hahn and Thorndike (46) used eight experimental groups varying in size, when approximate equivalence had been secured, from six to nineteen fourth-, fifth-, sixth-, and seventh-grade pupils. These groups were subjected to ninety minutes of practice in addition, divided into periods of 5, $7\frac{1}{2}$, 10, $11\frac{1}{4}$, 15, 20, and 22 minutes' duration. While the use of the practice sheets would seem to make

negligible the teacher factors, it is possible that an important extra-school factor was uncontrolled. The investigators state:

It should be kept in mind throughout the reading of what follows that any child was free to write out sums and to practice with them at home, during the course of the experiment no attempts were made to prevent practice apart from the specified practice in school.

The differences favor, but not significantly, the longer practice intervals. More dependence could be placed on this conclusion if larger groups had been used and used with more adequate control of non-experimental factors.

Reed (102) used four groups of 60, 50, 51, and 42 first- and second-year college students. The scores on the initial test in addition indicate that these groups were only approximately equivalent. One group practiced addition for a period of one hour, while the other groups practiced an equal amount of time distributed in periods of twenty minutes a day for three days, ten minutes a day for six days, or ten minutes twice a week for six weeks. It should be mentioned that the initial ten minutes of practice and the final nineteen minutes constituted the initial and final tests. The results favor significantly the distributed practice as compared with the one hour non-distributed practice. With respect to the distributed practice, the results favor, but not significantly, the daily twenty-minute practice periods. The chief criticisms of this experiment are that it was conducted with adults and that the groups having the distributed practice were initially superior. Hence, its conclusions are probably not applicable to school children. The adults but relearned an old skill. Results might be quite different with new learning.

Norem and Knight (91) used twenty-five third-grade pupils in their investigation of the distribution of practice effective for retention or maintenance of skill in multiplication. The parents of the pupils were requested to refrain from assisting them in drill at home, and the pupils were instructed not to practice except when required to do so by the experiment. After an initial administration of two tests, given a week apart, which disclosed unlearned combinations, each pupil was individually drilled to the point of mastery of his formerly unlearned combinations. The pupil was then tested once a week for a period of six weeks on these newly mastered combinations, and then once a month for three months. The analysis of the practice and test achievements of these twenty-five pupils is a commendable feature of this experiment. It would seem to justify the conclusion that one practice a week is sufficient for maintenance of skill in multiplication after mastery has been attained, so far as this group of pupils

is concerned. It is probable that this investigation should be repeated with larger groups for greater reliability in the findings.

3. Justified conclusions. The conclusions of Kirby (57) and of Hahn and Thorndike (46) are opposed to each other, while that of Reed (102) tends to agree with that of Hahn and Thorndike (46). The conclusions of Norem and Knight (91) seem reliable for the pupils used in their experiment, but do not seem more than suggestive for pupils in general. The conflicting testimony, plus the obvious faulty techniques of the experiments, prevents the authors from stating a justified conclusion.

It would seem, however, that until more adequate experimental evidence has been presented, the teacher will be acting wisely in employing intervals approximately twenty minutes in length with a frequency of one a day until mastery has been attained. After this objective has been reached, shorter practice periods distributed at longer intervals will possibly serve to maintain skill.

THE INFLUENCE OF REQUESTS FOR SPEED OR ACCURACY ON ACHIEVEMENT IN THE FUNDAMENTALS

1. Summary of reported conclusions. The influence of requests for speed or accuracy has been studied in three experiments.³² Wimmer (124) has reported that "the difference in progress made by the two groups, one being drilled for accuracy and the other for speed is not very large." Messick³³ reports that if speed is the objective of achievement in addition, it makes little difference which is requested, speed or accuracy. However, if accuracy is the objective, it is much better to request accuracy rather than speed. He states, "In teaching addition to pupils of the fourth and fifth grades of the elementary schools it is better to emphasize accuracy rather than speed." Myers³⁴ concludes that requests for speed are causes of inaccuracy in the fundamentals. "One may conclude that the loss to learning efficiency from the strong speed pressure as applied to the simple number combinations in arithmetic under which many school children must work in school today is appalling."

³²There have been several investigations of the relation of speed to accuracy in the fundamentals of arithmetic; see:

Bird, G. E. "A Test of Some Standard Tests," *Journal of Educational Psychology*, 11:275-83, May, 1920.

Courtis, S. A. "Courtis Standard Research Tests: Third, Fourth, and Fifth Annual Accountings, 1913-16," *Bulletin* No. 4. Detroit: Department of Cooperative Research, 1916. 112 p.

Luderman, W. W. "Speed and Scholarship Arithmetical Accuracy," *School Science and Mathematics*, 25:522-24, May, 1925.

Monroe, W. S. "A Report of the Use of the Courtis Standard Research Tests in Arithmetic in Twenty-Four Cities," *Studies by the Bureau of Educational Measurements and Standards*, No. 4. Emporia: Kansas State Normal School, 1915. 94 p.

Phelps, C. L. "A Study of Errors in Tests of Adding Ability," *Elementary School Teacher*, 14:29-39, September, 1913.

³³Messick, A. I. "Effect of Certain Types of Speed Drills in Arithmetic," *Mathematics Teacher*, 19:104-09, February, 1926. (75)

³⁴Myers, G. C. "The Price of Speed Pressure in the Learning of Number," *Educational Research Bulletin* (Ohio State University), 7:265-68, September 19, 1928. (84)

2. Evaluation of the experiments. The experiment of Wimmer (124) was described and evaluated on page 32. His conclusion with respect to speed versus accuracy may not be regarded as dependable. Messick (75) used two groups of 136 fourth- and fifth-grade children. No attempt was made to secure equivalence. One group practiced addition four minutes a day for twenty days, with emphasis on speed. The other group practiced addition for the same length of time, but requests were made for accuracy rather than for speed. The final tests revealed a certainly "statistically" significant difference in accuracy in favor of the group for which accuracy was emphasized. The small difference in speed also in favor of this group cannot be regarded as "statistically" significant. This experiment is faulty in that no attempt was made to secure equivalence. There is some reason for believing that important non-experimental factors were not adequately controlled. The experiment was rather short in duration.

Myers (84) used one group of ten first-grade children. These children, who had been practiced for two months in addition, were administered a test, the results of which indicated almost 100 per cent accuracy. After two years, "The ten who were still in school were studied again. In the meantime, these children . . . had been exposed to rapid-fire drills in the simple addition facts and the basic subtraction facts. The test-flash card . . . was their torturer almost daily . . . They were frequently subjected to games in which the fastest answers won." The children were then subjected to five practice-test periods, after each of which they were told that they had done very well and were urged to go faster. The decrease in accuracy as more and more emphasis was placed on speed is significantly shown in this experiment. Myers is to be commended for prolonging his investigation over so long a period of time. He is to be criticized for securing data from so small a group, for failure to employ a control group, and for creating what appear to be abnormal conditions. It is possible that the conditions to which these children were subjected are not typical of good, or even usual, school practice.

3. Justified conclusions. While dependable conclusions must await further controlled experimentation, it seems justifiable to recommend requests for accuracy rather than requests for speed. In any case, it seems justifiable to hold that requests for accuracy should precede requests for speed. After pupils have attained satisfactory accuracy on a given level of difficulty, a teacher is possibly justified in encouraging them to increase their rate.

CHAPTER IV

METHODS OF TEACHING PUPILS TO SOLVE VERBAL PROBLEMS

It is commonly assumed that the responses made by pupils when presented with verbal problems in arithmetic are the result of reflective thinking. Consideration is given in the first part of this chapter to investigations of the nature of pupil responses to verbal problems. The experimental factors of the experiments summarized in the second part of the chapter are variations in types of verbal problems and of problem statements, and those in the third and final portion of the chapter are various methods of teaching pupils to solve verbal problems in arithmetic.

THE NATURE OF PUPIL RESPONSES TO VERBAL PROBLEMS

1. **Summary of reported conclusions.** Three studies have been reported on the problem of the part played by reasoning when pupils attempt to solve verbal problems in arithmetic. Bradford¹ reported from an analysis of test results that "arithmetical work is not done in a critical frame of mind." This conclusion has since been substantiated by the more comprehensive investigation of Monroe,² in which the conclusion was reached that "a large per cent of seventh-grade pupils do not reason in attempting to solve arithmetic problems Many of them appear to perform almost random calculations upon the numbers given. When they do solve a problem correctly, the response seems to be determined largely by habit." Kline and Anderson³ have reported a laboratory study, the findings of which indicate the nature of the dual role of specific habits and reasoning abilities in solving verbal problems in arithmetic.

2. **Evaluation of the investigations.** The data in the investigations of both Bradford (11) and Monroe (79) were collected by means of a single administration of tests. The tests of Bradford (11), which were administered to several hundred pupils in Standards VII and VIII in certain elementary schools in England, were composed of examples impossible of solution, of which the following quoted from the report are illustrative:

¹Bradford, E. J. G. "Suggestion, Reasoning, and Arithmetic," *Forum of Education*, 3:3-12, February, 1925. (11)

²Monroe, W. S. "How Pupils Solve Problems in Arithmetic," *University of Illinois Bulletin*, Vol. 26, No. 23, Bureau of Educational Research Bulletin No. 44. Urbana: University of Illinois, 1929. 31 p. (79)

³Kline, L. W. and Anderson, P. K. "The Role of Habit in Reasoning," *School Science and Mathematics*, 26:156-67, February, 1926. (59)

1. If the distance from Arles to St. Brieuc is 500 miles, and from Vire to St. Malo is 50 miles, how far is it from St. Brieuc to St. Malo?
2. If Henry VIII had six wives, how many had Henry II?

The extent to which attempts were made to solve such problems was taken by Bradford to be indicative of the absence of critical reflective thinking in the solving of arithmetical problems by school children. While this conclusion seems reasonably dependable, it should be remembered that the data refer to the children of English schools and for this reason may be somewhat less applicable to American children. It is in agreement, however, with the conclusion of the investigation reported by Monroe (79).

Monroe (79) secured his data by administering a test to 775 sixth-grade, 5902 seventh-grade, and 2579 eighth-grade pupils in forty-one Illinois cities. These pupils were divided into four groups, and equivalence was secured by distributing the tests to the pupils in a random manner.

In order that each of the tests might be given to a random sample of pupils, the four tests were arranged in alternate order so that when distributed to the pupils in the class, the first, fifth, ninth, thirteenth, and so forth, would receive Test A; the second, sixth, tenth, fourteenth, and so forth, would receive Test B; the third, seventh, eleventh, fifteenth, and so forth, would receive Test C; the fourth, eighth, twelfth, sixteenth, and so forth, would receive Test D. Since the tests were to be given in a large number of classes, it seemed that this plan of sampling would provide equivalent groups

It is evident that the four groups were equivalent not only in arithmetical ability but also with respect to teachers, textbooks, and other factors. In general each of the four equivalent groups was equally represented in each classroom, and this representation was secured in a random fashion.

The tests administered to these groups differed only in the terminology used in stating the problems. For example, in Test A, the second problem is stated in simple terminology, all of the data given are relevant, and the setting is concrete. In Test B, technical terminology is used, all the data given are relevant, and the setting is concrete. The difference in the statement of the problem in these two tests is the change from simple terminology to technical terminology. In Test C, the problem is stated in simple terminology, the data given are relevant, and the setting is abstract. In Test D, technical terminology is used, irrelevant data are included, and the setting is abstract. The problems of the tests are so stated that comparisons are possible with respect to the relative influences on correctness of response of simple and technical terminology, wholly relevant data and data partially irrelevant, and concrete and abstract setting. These com-

parisons are made for the data of this investigation, and the results are presented in tabular form in the report of the research.

The techniques used in this study appear to be reasonably free from criticism. There seems to be little question that the sample of pupils was representative, and the groups used, equivalent with respect to all significant factors. The data secured seem to be of sufficient quality to warrant the statement that responses of pupils to verbal problems are usually characterized by absence of reasoning.

The experiment of Kline and Anderson (59) was conducted with four adults in a psychological laboratory. Time and accuracy were recorded for the responses to four hundred questions, such as "If Thursday is the twelfth, what day is the eighteenth?" The conclusions of this experiment are interesting, but may not safely be applied to school children. It would seem, however, that Kline and Anderson have made but another attempt to prove the obvious. It is commonly recognized that there is close interdependence between specific habits and reasoning.

3. Justified conclusions. The data secured in these three investigations appear to justify the conclusions stated, insofar as they apply to the groups of pupils to which the tests were given and by which the test exercises were used. The generalization of the conclusions may be questioned, especially for all types of problems and for all conditions of responding to them. Hence, the generalization should be considered tentative. It should also be noted that these investigations deal with the question of what responses pupils make as the result of the instruction they have received. They do not consider the type of responses that pupils should make.

THE EFFECT OF DIFFERENT TYPES OF PROBLEMS AND PROBLEM STATEMENTS

1. Summary of reported conclusions. Myers,⁴ Hydle and Clapp,⁵ Washburne and Morphett,⁶ Bowman,⁷ Mitchell,⁸ Monroe,⁹ Wheat,¹⁰ and Osburn and Drennan¹¹ have reported conclusions relative to the

⁴Myers, G. C. "Imagination in Arithmetic," *Journal of Education*, 105:662-63, June 13, 1927. (83)
⁵Hydle, L. L. and Clapp, F. L. "Elements of Difficulty in the Interpretation of Concrete Problems in Arithmetic," *Bureau of Educational Research Bulletin* No. 9. Madison: University of Wisconsin, 1927. 84 p. (50)

⁶Washburne, C. W. and Morphett, M. V. "Unfamiliar Situations as a Difficulty in Solving Arithmetic Problems," *Journal of Educational Research*, 18:220-24, October, 1928. (118)

⁷Bowman, H. L. "The Relation of Reported Preference to Performance in Problem Solving," *University of Missouri Bulletin*, Vol. 30, No. 36, Education Series, No. 29. Columbia: University of Missouri, 1929. 52 p. (10)

⁸Mitchell, Claude. "The Specific Type of Problem in Arithmetic versus the General Type of Problem," *Elementary School Journal*, 29:594-96, April, 1929. (76)

⁹Monroe, *op. cit.*

¹⁰Wheat, H. G. "The Relative Merits of Conventional and Imaginative Types of Problems in Arithmetic," *Teachers College, Columbia University Contributions to Education*, No. 359. New York: Bureau of Publications, Teachers College, Columbia University, 1929. 124 p. (121)

¹¹Osburn, W. J. and Drennan, L. J. "Problem Solving in Arithmetic," *Educational Research Bulletin* (Ohio State University), 10:123-28, March 4, 1931. (95)

effect upon pupil responses of certain variations in the statement of the problems. Myers (83) administered two problems to fifth-grade pupils and reported that these pupils were able to solve the "imaginatively stated" one much more easily. Hyde and Clapp (50) studied the following characteristics of arithmetical problems in an effort to determine whether or not these characteristics were causes of difficulty in problem solving:

1. Objective setting
2. Size of numbers
3. Unfamiliar objects
4. Arrangement in a series
5. Nonessential elements
6. Visualization vs. experience
7. Project vs. problem form of statement
8. Symbolic terms

Variations of these characteristics, with the exception of the arrangement of similar problems in series and the presence of non-essential elements were found to be "statistically" significant causes of difficulty. In addition to this conclusion the authors state that problem solving for pupils is largely a matter of visualization. Problems should be formulated with this in mind in the earlier stages of learning, but in order that generalizing ability might be engendered, it is concluded that the pupils should have as learning exercises a considerable number of problems not related to their first-hand experiences.

Washburne and Morphett (118) report that fifth-grade pupils achieve better results with familiar problems than with those containing unfamiliar elements. The following problems quoted from the report are illustrative of those used in his study; the first is in unfamiliar terminology, and the second, in familiar terminology:

A merchant sold 20 bags of charcoal. Each bag held 35 pieces. How many pieces of charcoal did he sell?

The girls have to make 30 boxes of taffy. Each of the boxes holds 25 pieces. How many pieces of taffy do they have to make?

Bowman (10) reported that pupils of high ability, as measured in his study, performed equally well on the following types of problems:

1. Problems based upon adult activities
2. Problems based upon children's activities
3. Problems whose setting is in the field of science
4. Problems so stated as to take on the nature of a puzzle
5. Problems of pure computation only, where directions for the right procedure are given

Pupils of lower ability showed a higher relative degree of performance on problems of the pure computation type. Mitchell (76) re-

ported that "Problems with definitely expressed numerical quantities seem to be more readily understood and solved than problems of a general nature involving general principles." The following examples illustrate the types of problems compared in this study. The first is a specific problem, and the second, a general problem.

The width of a room is 10 feet, and its length is 15 feet. Find its perimeter.

If you know the length and the width of a room, how can you find the perimeter?

Monroe (79) reported as another of the conclusions of his study that "If the problem is stated in the terminology with which they [the pupils] are familiar and if there are no irrelevant data, their response is likely to be correct." Wheat (121) determined the relative achievements of pupils with conventionally-stated problems and imaginatively-stated problems. He reported that differences in achievement are negligible. The first of the examples quoted below illustrates the conventional type of statement; the second of the examples illustrates the imaginative type.

Margaret spent \$3.68 for handkerchiefs at 23 cents each and gave one-fourth of them to her sister. How many did her sister get?

Margaret had been shopping all morning for Christmas presents. She had bought presents for her father and mother and brothers but could not decide what to get for her sister and several of her friends—there were so many things to pick from. Just then she saw some pretty handkerchiefs which were marked 23 cents each. These were just what she wanted, so she counted her money, found she had \$3.68, and spent all of it for handkerchiefs. She kept out one-fourth of the handkerchiefs to give to her sister and gave the rest to her friends. How many did she keep out to give to her sister?

Osburn and Drennan (95) have reported a recent experiment in which vocabulary difficulty did not appear to be a significant factor in problem-solving achievement. These investigators conclude that their data "seem to indicate that pupils are able to sense the meaning of problems even if they do not understand all the words." The conclusion is also reported that a few of the most important problem types should be taught thoroughly, with the expectation that transfer of training will take care of the remainder.

2. Evaluation of the experiments. Myers (83) administered his two problems to 513 fifth-grade children. One hundred and ninety-seven solved the first problem correctly, while 253 correctly solved the second and more imaginatively-stated problem. It would seem, probably, that the difference is due to practice effect rather than to the fact that the second problem was more imaginatively stated than the first.

Hydle and Clapp (50) constructed tests in which the problems were paired with respect to each of the elements of difficulty investi-

gated. That is to say, a problem appearing in one form of the test differed from its mate in the other form with respect to a given element. For example, in the case of symbolic terms one problem statement would contain symbols, such as X, Y, and Z, instead of the names of objects given in the other problem statement. The tests included five pairs of problems for each of the following elements of difficulty: (1) objective setting, (2) size of numbers, (3) unfamiliar objects, (4) arrangement in a series, (5) nonessential elements, (6) visualization vs. experience, (7) project vs. problem form of statement, (8) symbolic terms. The tests were administered to pupils varying in number from 5870 to 7029. These pupils were widely distributed in village and city schools. Those taking the tests were divided into two groups of approximately equal ability as shown by a test of twenty-five problems of a concrete character. The statistical interpretation of the data indicated that variations in six of the eight elements investigated might dependably be expected to cause difficulty.* These elements are (1) objective setting, (2) size of numbers, (3) unfamiliar objects, (4) visualization vs. experience, (5) project vs. problem form of statement, and (6) symbolic terms. Hyde and Clapp are to be commended for their comprehensive and intensive investigation. The possible invalidity of their problem tests is adequately recognized in the report of the study. The investigators are to be commended for this and, in the opinion of the present writers, for *not* contending that the arithmetic curriculum should be so constructed that difficult elements in problem solving be eliminated.

Washburne and Morphett (118) used a single group of 441 fifth-grade pupils in six different towns. A test of eight pairs¹² of problems was administered to all of these children. The results appear to be "statistically" significant in favor of the problems containing familiar elements. The data collected would seem to be sufficiently reliable to warrant acceptance of the conclusion. However, this experiment would seem to be but another attempt to prove the obvious. A more worth while investigation would be one that would attempt to show whether or not problems containing unfamiliar elements should be used as learning exercises.

Bowman (10) administered both forms of his test to a total of 564 seventh-, eighth-, and ninth-grade pupils of Sedalia, Missouri. Evidence is presented to show that the pupils of this group represent an approximately normal distribution of intelligence and are typical of the grades they represent with respect to parentage, parental occu-

¹²An illustration of one of the pairs of problems is given on page 53.

pations, and environment.¹³ Each of the two test forms contained twenty-five problems of the types previously referred to. At the bottom of each page of the forms was placed the following statement to be completed by the pupil: "The problem on this page I liked best is No. —." This was done to secure data relevant to preferences for different types of problems.¹⁴ The coefficients of reliability and of validity for the test as a whole were quite high. The coefficient of reliability was reported as $.95 \pm .003$ in the measurement of performance and $.77 \pm .01$ in the measurement of preference, and the coefficient of validity was reported as $.82 \pm .01$ when the scores secured from an administration of the Stanford Arithmetic Reasoning Test were used as the criterion. The representativeness of the group and the comparatively high reliability and validity of the instrument used constitute strong arguments for the dependability of the conclusions that pupils of high ability perform equally well on (1) problems based upon adult activities, (2) problems based upon children's activities, (3) problems whose setting is in the field of science, (4) problems so stated as to take on the nature of a puzzle, (5) problems of pure computation only, and that pupils of lower ability perform relatively better on problems of the purely computational type.

Mitchell (76) administered a test containing fifteen quantitative problems and fifteen general problems—problems without expressions of numerical quantities—to seventy eighth-grade and sixty seventh-grade pupils. The mean difference in scores between the two types of problems is sufficiently large to seem to be "statistically" significant, although no standard or probable error is reported. The dependability of the findings may be questioned, however, because of certain faults in the data. The sample of pupils is too small to be regarded as representative. It may be that the pupils had greater difficulty with the general, or non-quantitative, type of problem because of lack of experience with problems of this type.

Wheat (121) administered tests containing ten pairs of conventional and imaginative problems to approximately two thousand fifth-, sixth-, and eighth-grade pupils in several towns in different parts of the country. The differences in achievement between the conventional and imaginative types of problems were not of sufficient magnitude to be considered "statistically" significant, with the possible exception that the conventional type of problem required much less time. Wheat is to be commended for the size and representativeness of his sample, but his procedures for handling and inter-

¹³While measures of intelligence of some of the pupils are not reported, there is no reason to believe that they were less typical of children in general than those for whom data are reported.

¹⁴This matter will be referred to again in the summary of research on motivation of learning in arithmetic. See page 81.

preting his data have been seriously criticized. Osburn¹⁵ states that Pearson coefficients of correlation are computed from unsuitable data:

In at least two cases correlations are figured which are partly based upon the number of problems solved. The distribution of the number of problems solved is not normal; in fact it is clearly of the U type. The use of the Pearson coefficient of correlation with distributions of this sort may be justifiable if the regression lines are rectilinear. This necessary condition is not substantiated, and the use of the Pearson technique is therefore open to question.

Again, the Pearson correlation was originally intended for use with two variables only. In a number of cases in this study it is used where three and even four variables are involved. For example, a correlation is shown between intelligence quotients and indices of similarity scores. In this case four variables are really involved, but they appear as two because quotients of respective pairs are used. . . . This is handy, but hardly justifiable, as a statistical procedure.

Osburn also criticizes the study from other points of view. He states, "In conventional problems, as here defined, the setting is left to the imagination, while in the imaginative problems the setting is made explicit by description but is still not perceptually present." The critic points out that the pupils quite possibly received previous training only on the conventional type of problem.

In spite of the fact that they had had little or no training in the solution of imaginative problems the pupils did well with them. This might mean the existence of transfer, or it might indicate a marked advantage for the imaginative type when the factor of previous training is properly controlled by acceptable scientific techniques.

Finally, Osburn contends that Wheat is to be criticized for assuming that arithmetic material should be used which can be bought cheaply and taught quickly and easily. Osburn holds that the objectives of arithmetic must be considered here. "The question therefore is not which problem is most economical to teach, or to buy, but which one will better prepare the pupil for quantitative thinking in real conditions—the sorts of situations which he will meet in life." Osburn then presents arguments for the imaginative type of problem.

The present writers are inclined to grant that most of Osburn's criticisms appear to be justified. It should be pointed out, however, that Osburn is somewhat inconsistent. For example, he holds that the two types of problems are synonymous and then contends that training has been different with respect to each. If they are synonymous, why should each not be equally well adapted to engender those abilities accepted as the objectives of arithmetic? After all, it would seem that the conclusion that "pupils of the intermediate grades are

¹⁵Osburn, W. J. "Two Recent Books on Arithmetic," *Educational Research Bulletin* (Ohio State University), 9:66-73, February 5, 1930.

neither hindered nor helped in their problem practice exercises by problems of the imaginative type, when no limits are imposed upon the amounts of time of the practice periods," may be accepted as fairly dependable until better evidence has been obtained experimentally which reverses it.

Osburn and Drennan (95) had teachers of two classes of third-grade children teach a representative list of problems with particular emphasis on the "cues," or language aspects, of the problems. An examination made up of twenty verbal problems containing new ones, but no additional vocabulary difficulty, was given after six weeks of such instruction. On the next day, another test was administered containing twenty problems which involved vocabulary difficulties, illustrated by such terms as narcissus, gypsum, tortoise, chemist, sulfuric acid, and excavating. The data indicate that the pupils made very acceptable scores on both tests. The investigators suggest that the changes in vocabulary may have been a factor of little significance, because "mainly just 'nouns' were changed, and since the test was given the next day after the first test, that the pupils sensed the similarity of Test II to the test of the day before." This appears to be a very serious limitation of this investigation. The present writers are inclined, therefore, to give little weight to the conclusions of other studies of the influence of terminology on problem-solving achievement in arithmetic.

3. Justified conclusions. These eight studies of the effect of different types of problems and problem statements are not comparable, and, hence, it is difficult to synthesize the findings. Most of them, however, support the principle that pupils make higher scores on tests consisting of familiar problems, or problems stated in familiar terminology. The conclusion that pupils respond more correctly to problems stated in concrete rather than imaginative or abstract form, with irrelevant elements excluded, and related to activities experienced by children is less unanimously supported by the experimental evidence. This generalization is an obvious inference from the psychology of learning, but these studies contribute to our understanding of what makes a problem unfamiliar.

METHODS OF TEACHING PUPILS TO SOLVE VERBAL PROBLEMS

1. Summary of reported conclusions. Newcomb,¹⁶ Stevenson,¹⁷ Greene,¹⁸ Clark and Vincent,¹⁹ Washburne and Osborne,²⁰ Lutes,²¹

¹⁶Newcomb, R. S. "Teaching Pupils How to Solve Problems in Arithmetic," *Elementary School Journal*, 23:183-89, November, 1922. (90)

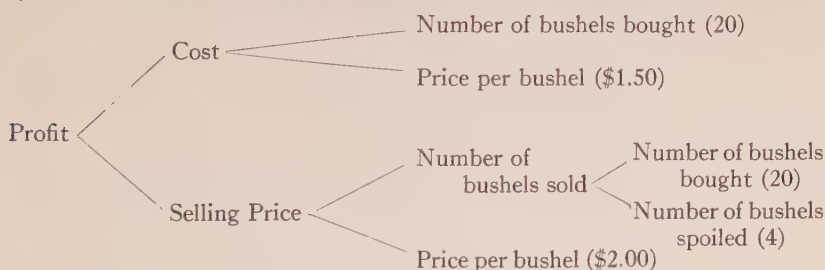
¹⁷Stevenson, P. R. "Increasing the Ability of Pupils to Solve Arithmetic Problems," *Educational Research Bulletin* (Ohio State University), 3:267-70, October 15, 1924. (112)

¹⁸Greene, H. A. "Directed Drill in the Comprehension of Verbal Problems in Arithmetic," *Journal of Educational Research*, 11:33-40, January, 1925. (42)

Washburne,²² Hanna,²³ and Adams²⁴ have reported studies on methods of teaching pupils how to solve problems in arithmetic. Newcomb (90) concluded that the pupils in his experiment who were supplied with sheets of general directions for solving verbal problems achieved more, particularly with respect to speed, than the pupils not so supplied. Stevenson (112) secured effective results with a large group of pupils who were taught to read and analyze problems by the provision of systematic training in finding the facts pertaining to the problem, in deciding upon the processes to be used, and in finding the answer in round numbers. Greene (42) reported that training in selecting and recognizing the process involved in the solution of a problem is more effective in securing correct solutions from pupils than when such training is not given. He states in this connection, however, that "This drill, strangely enough, seems to increase the accuracy of problem solution more than the ability to select the correct principle in solving the problem"

Clark and Vincent (26) compared the relative effectiveness of the conventional and graphical methods of solving verbal problems in arithmetic. The results are favorable, but not significantly so, to the graphical method. This method is illustrated by the following example quoted from the report:

A grocer bought 24 bushels of potatoes at \$1.50 per bushel. Four bushels spoiled. The others were sold at \$2.00 per bushel. Find his profit.



The pupil is directed to think of the diagram as illustrating the following: To find the profit, I should have to know the *cost* and the *selling price*; to find the cost I would have to know *the number of bushels bought* and *the price per bushel*;

¹⁹Clark, J. R. and Vincent, E. L. "A Comparison of Two Methods of Arithmetic Problem Analysis," *Mathematics Teacher*, 18:226-33, April, 1925. (26)

²⁰Washburne, C. W. and Osborne, Raymond. "Solving Arithmetic Problems," *Elementary School Journal*, 27:219-26, 296-304; November, December, 1926. (119)

²¹Lutes, O. S. "An Evaluation of Three Techniques for Improving Ability to Solve Arithmetic Problems: A Study in the Psychology of Problem Solving," *University of Iowa Monographs in Education*, Series 1, No. 6. Iowa City: University of Iowa, 1926. 42 p. (68)

²²Washburne, C. W. "Comparison of Two Methods of Teaching Pupils to Apply the Mechanics of Arithmetic to the Solution of Problems," *Elementary School Journal*, 27:758-67, June, 1927. (117)

²³Hanna, P. R. *Arithmetic Problem Solving: A Study of the Relative Effectiveness of Three Methods of Problem Solving*. New York: Bureau of Publications, Teachers College, Columbia University, 1929. 68 p. (47)

²⁴Adams, R. E. *A Study of the Comparative Value of Two Methods of Improving Problem Solving Ability in Arithmetic*, Philadelphia: University of Pennsylvania, 1930. 68 p. (1)

to find the selling price I would have to know the *number of bushels sold* and the *price per bushel*.

Washburne and Osborne (119) compared the relative effectiveness of three methods of teaching pupils to solve verbal problems in arithmetic.

Method 1 is to train children in the solving of problems by giving them a large number of problems—no special technique

Method 2 is to train children to *analyze* problems. It is a definite technique of attacking problems.

Method 3 is to train children to see the *analogy* or similarity between difficult written problems and corresponding easy oral problems and thereby to decide what process to use in attacking the difficult problems.

They state in their conclusions:

Training in the seeing of analogies appears to be equal or slightly superior to training in formal analysis or the superior half of the children; analysis appears to be decidedly superior to analogy for the lower half; *but merely giving many problems, without any special technique of analysis or the seeing of analogies, appears to be decidedly the most effective method of all.*

Lutes (68) compared the relative effectiveness of (1) drilling pupils in computation only (2) drilling pupils in choosing operations, (3) drilling pupils in choosing correct solutions, along with emphasis on reading problems correctly, and (4) the traditional method of teaching pupils to solve verbal problems. The results are significantly in favor of drilling pupils in computation. Washburne (117) compared the achievement of pupils who were taught the fundamental processes *as applied* to verbal problems with the achievement of pupils who were taught fundamental processes and verbal problems separately. The results were not significantly in favor of either method.

Hanna (47) compared the relative effectiveness of the dependencies method (graphic or diagrammatical), the conventional-formula (four steps) method, and the individual, or informal, method of teaching pupils to solve arithmetical problems. The dependencies method is similar to the graphic method of Clark and Vincent (26). The conventional formula method consists of the following steps:

1. What is asked for in the problem?
2. What is given in the problem?
3. How should these facts be used to secure the answer?
4. What is the answer?

In the individual method the pupils were allowed to use any method of problem analysis which they desired. The conclusions of this study are distinctly unfavorable to the conventional-formula method. The dependencies and the individual methods were found to be approximately equal in effectiveness.

Adams (1) compared the relative effectiveness of teaching pupils to solve verbal problems in arithmetic by an analytical method and by one in which no attempt at analysis was made. The analytical method is illustrated by the following quotation concerning a demonstration of the solution of a one-step problem by the teacher:

How many apples will Tom need to fill 4 baskets if he puts 6 apples into each basket?

1. The problem is read.
2. "What are we asked to find?"
3. "What do we know that will help us to find the answer?"—that there are 4 baskets and that Tom puts 6 apples into each.
4. "What will be the name of the answer?"—apples
5. "Will he need more or less than 6 apples?" Select the number in the problem that corresponds to the name of the answer. This device cannot be used in some division problems.
6. "What two operations give us more for an answer?"—addition and multiplication
7. "Which shall we use here?"—multiplication
"Why could we not use addition?"—because you cannot add "apples" and "baskets."

The non-analytical method of teaching the solution of verbal problems is illustrated in the quotation below:

How much will Frank have to pay for 3 cans of peas that are sold for 18 cents a can?

1. Read the problem carefully.
2. Teacher asks, "What are we asked to find?"
3. "What do we know that will help us find it?"
4. "Shall we add, subtract, multiply, or divide?"
5. The solution is then performed.

The conclusions reported by Adams are favorable to the analytical method when used with third-grade children. The evidence does not significantly favor either method for fourth-grade children. Adams states in this connection that possibly insufficient time was devoted to the experiment to permit the breaking down of problem-solving habits previously learned.

2. Evaluation of the experiments. Newcomb (90) used four experimental and two control groups varying in size from fourteen to thirty-six pupils each. These groups, which were made up of seventh- and eighth-grade pupils, were approximately equivalent in arithmetical reasoning ability as shown by the Stone Reasoning Test. The experimental groups were taught one problem a day for twenty days, by means of sheets of general directions for solving verbal problems, while the control pupils were taught the same problems in the traditional fashion. At the end of the experimental period of twenty days the Stone Reasoning Test was again administered. The results

showed that the pupils who had used the sheets of general directions were significantly better in speed, but only slightly better in accuracy. The experiment may be criticized from several standpoints. The groups used cannot be said to be representative of seventh- and eighth-grade pupils in general, nor do they appear to have been sufficiently equivalent. No mention is made of any attempt to control important non-experimental factors. One suspects that the experimental method was applied with greater zeal. Newcomb is justified, however, in expressing his conclusion in favor of the sheets of general directions with appropriate limitations.

Stevenson (112) used a single group of 1027 fifth-, sixth-, and seventh-grade pupils in eight localities. These pupils were taught to read and analyze problems and to estimate answers in round numbers for a period of twelve weeks. The gains in achievement are certainly significant. While Stevenson shows that this method is effective, he does not show that it is more effective than other methods. It is unfortunate that control groups were not used.

Greene (42) used an experimental group of sixty-two pupils and a control group of thirty pupils. These pupils were all in the sixth grade and were attending four schools in one system. The groups were not equivalent in arithmetical reasoning ability, as was shown by the Monroe test. The pupils in the experimental group were given training in recognizing and selecting the process involved in the solution of the problem, while the control pupils did not have the advantage of such instruction. Both groups were practiced ten minutes a day for eight days, at the end of which time the Monroe Standard Reasoning Test was administered again. The investigation sought to correct for lack of equivalence by correcting the gain of one of the groups by proportion, a procedure that may not be sanctioned, unless it is proved that practice has no effect on individual differences. The experiment is to be further criticized for the failure of the investigator to continue it for a sufficient length of time to reveal significant differences in achievement. It should be mentioned that the conclusions favorable to the instructional method used with the experimental group are expressed with appropriate restrictions.

Clark and Vincent (26) used two groups of forty seventh- and eighth-grade pupils each in one school. These groups were equated with respect to intelligence as measured by the Stanford Revision. One group was taught by the conventional method for seven recitations, while the other group was taught by the graphic method.²⁵ At the close of the experiment the relative achievement of the two

²⁵See page 59 for an illustration of the "graphic" method.

groups was measured by the arithmetic section of the Stanford Achievement Test, Form A. The results appear to be somewhat significantly in favor of the graphic method. These experimenters are to be commended for their care in securing equivalence and for the precision with which they describe the compared factors in the report of their research. They are to be criticized for the short duration of their experiment and for failure to mention the use of procedures to secure control of important non-experimental factors. One wonders whether the graphic method advocated by them would engender abilities compatible with recognized arithmetical objectives. In the opinion of the present writers, it might be responsible for the engendering of habits which will later need to be unlearned.

Washburne and Osborne (119) used three groups of sixth- and seventh-grade children in eighteen schools, in investigating the relative effectiveness of (1) assigning large numbers of problems—no special technique, (2) training in analysis of problems, and (3) training in seeing analogies between difficult written problems and easy oral problems. These groups were of the following sizes: 322, 307, and 134 pupils. Equivalence was sought with respect to (1) problem-solving ability, (2) ability with fundamental processes, (3) intelligence, (4) chronological age, and (5) judgments of teachers with respect to capacity. The following quotation, from the directions issued to the participating teachers, indicates the precautions taken to control important non-experimental factors:

All other factors should, therefore, be made equal except for the particular differences in method which constitute the experiment. To this end, the same teacher teaches both groups. She does not know the children in one group better than she knows those in the other. The children who are taught earlier in the day one week should change class periods with the others the next week. The amount of time spent by the two groups should be the same. The amount of time, if any, given to drill in the fundamental processes will be the same and the method the same. The amount of oral work, or work done by the class with the teacher, will be the same. No home work will be permitted. No extra time will be allowed in school with this exception: children who have been absent may make up in school the number of periods they missed and do the problems they missed. If this is done, it must, of course, be done in both groups.

The experiment continued for six weeks, at the end of which time a specially devised problem test was administered. It is unfortunate that the experimenters did not report measures of the "statistical" significance of the differences in achievement. They do not appear to be of sufficient magnitude to be significantly in favor of any one of the methods. While many of the techniques used in this experiment were excellent, one wonders whether it is not somewhat futile to com-

pare methods, each of which contains some logically excellent characteristics.

Lutes (68) used four groups of sixth-grade pupils in twelve elementary schools of Des Moines, Iowa. The following evidence is cited by the investigator relative to the representative character of the groups.

The twelve schools were scattered widely over the city in such a way as to include groups which were representative of widely diverse elements of the population, a wide range of native intelligence, of social status, and of personality of the teachers involved.

These groups, which varied in size from sixty to seventy-four pupils, were approximately equivalent with respect to arithmetical ability as measured by the Stanford Achievement Tests, Parts 4 and 5, and with respect to intelligence as measured by Scale A, Form 1, of the National Intelligence Test. The pupils in the first group were drilled in computation, those in the second group were trained in choosing operations, those in the third group were taught to choose correct solutions and to read problems, while those in the fourth group were taught by the traditional method. Considerable care was exercised in the control of non-experimental factors:

The same days of the week were used by each group, the same length of recitation period, and the experimenter spent practically the same amount of time with each group and each teacher. No home study was required in any case . . . though of course it is impossible to be certain that some of the pupils did not practice the skills at home in order to make a good showing in the test.

At the end of twelve weeks the second form of the Stanford Achievement Test, Parts 4 and 5, was administered. The differences in gains appear to be significantly in favor of the group drilled in computation. While the techniques of this experiment are for the most part excellent, one may raise the same question that was raised with respect to the preceding experiment. Each of the methods appears to be logically desirable. Why should the relative effectiveness of computational drill, drill in choosing operations, and drill in choosing correct solutions along with emphasis on reading problems correctly be compared?

Washburne (117) used two groups of 175 second-grade pupils, two groups of 177 fourth-grade pupils, and two groups of 240 sixth- and seventh-grade pupils of sixteen cities of northern Illinois. Equivalence was sought with respect to the following traits: (1) problem-solving ability, (2) ability in arithmetic mechanics, (3) mental age, (4) chronological age, and (5) general ability to work as judged by the teacher. The pupils in one group were taught the fundamental

processes in connection with verbal problems, while the pupils in the other group were taught fundamental processes and verbal problems separately. Evidence is presented in the report of the experiment which indicates that considerable care was exercised in the control of important non-experimental factors. At the end of six weeks the final tests were administered. The difference in achievement was not significantly in favor of either method. Many of the techniques used in this experiment are very commendable. It would seem, however, that the tests used were more valid with respect to the group which had practiced verbal problems over the longer period of time. The group which learned the fundamental processes in connection with verbal problems had their practice in verbal problems distributed in a way considered to be more psychologically effective.

Hanna (47) used three groups of seventy-five fourth-grade pupils and three groups of eighty-four seventh-grade pupils in his attempt to determine the relative effectiveness of the dependencies (graphic or diagrammatical), of the conventional-formula (four steps), and of the individual, or informal, methods of teaching pupils to solve arithmetical problems. The groups were shown to be equivalent (for both grade levels) with respect to intelligence and initial arithmetical ability. The arithmetic tests, the same forms of which were used at the beginning and end of the experiment, were the new Stone Test in Arithmetic Reasoning and the Stanford Achievement Test in Arithmetic Reasoning (Form A, Test 5). The teachers were given detailed written directions for conducting the experimental instruction. The materials of instruction were also carefully prepared. The pupils were given practice sheets, and during the first seven days of the experimental period, they were requested to work the problems thereon with the help of instructions given by the teacher. On the eighth day and on alternate days, until the close of the experiment, the pupils worked the problems on the sheets independently of the teacher. The experiment lasted six weeks, or a total of twenty practice periods. At the end of this time the final tests were administered. In addition to the differences in mean gains, and the "statistical" significance of these differences, the investigator reports learning-curve data secured by scoring the practice sheets for the days in which the pupils worked independently.

Hanna is to be commended for the many excellent techniques employed in his experiment. It would seem that he has rather adequately defined his experimental factors, secured equivalent groups, controlled important non-experimental factors, and measured achievement. It would seem that the only important adverse criticism to be

made with respect to the techniques used in this experiment has to do with the somewhat artificial conditions necessitated by procedures employed to control non-experimental factors. It would appear, however, that some sacrifice of usual schoolroom conditions is justified if adequate control of non-experimental factors is thus attained. The conclusion that the dependencies, or graphic, method and the individual, or informal, method are superior to the conventional-formula method appears to be reasonably dependable. The conclusion that the dependencies method is not significantly better than the individual method also appears to be reasonably dependable.

In his first experiment Adams (1) taught 834 pupils by the "method of analysis," 772 pupils by the method prescribed in the Philadelphia Course of Study in Arithmetic, and 507 pupils "by the methods usual to the teachers in charge." The pupils participating in the experiment were located in the third and fourth grades of ten Philadelphia public schools selected in an effort to secure representativeness and control of school and extra-school factors. The experiment lasted for a period of eight weeks. The analysis of the data showed that while the scores of the experimental classes were highest in only one instance, the greatest gains in achievement were made in these classes. In the second experiment 1033 experimental and 1065 control pupils were used. All of the teachers participating in the experiment were paired according to their teaching ability as estimated by supervisors, and other steps were taken in an effort to secure control of important non-experimental factors. The final test was administered at the end of seven weeks. The analysis of the data thus secured was quite inconclusive with respect to the relative effectiveness of the methods compared.

In the third experiment 1938 experimental and 1836 control pupils were used. The ninety-six school classes participating in the experiment were paired on the basis of class medians on the initial arithmetic test. The investigator contends that since there is a high correlation between intelligence and the problem-solving ability measured by the initial test, the experimental and control group were probably equivalent in intelligence. The argument is advanced, and quite rightly, that the use of experimental and control groups of such great size very probably secures adequate equivalence with respect to pupil characteristics through the operation of chance. The pupils in the experimental group were taught to solve problems by an analytical method, while no attempt at analysis was made in the teaching of the control pupils. Only one of the two methods was taught in any one school or by any single teacher. Data are pre-

sented to show that the teachers were approximately equivalent with respect to training, experience, and after-school professional training. Data are also presented to show that supervision of the experimental and of the control teachers was approximately the same. The final test was administered at the end of eight weeks. The data secured in the third experiment were also quite inconclusive with respect to the relative effectiveness of the compared methods, although the method of analysis was shown to be slightly more effective in the third grade.

Adams is to be commended for the many excellent techniques used in his experiments. He should be criticized, however, for failure to conduct his experiments over a longer period of time.

3. Justified conclusions. Several of the studies in this group contribute evidence in support of the generalization that systematic and persistent training in a procedure for attacking verbal problems results in higher scores on problem tests. This generalization is a fairly obvious inference from the Law of Exercise and the supplementary Law of Intensity.

With respect to relative evaluation of comparable methods of teaching pupils to solve problems, the findings are probably not highly dependable. It was pointed out in the evaluation of several of the experiments that the non-experimental factors of the zeal and skill of the teacher were inadequately controlled and that differences favoring a given method are possibly more justifiably attributable to these influences than to any merits inherent in the method. It may be concluded, therefore, that several methods of teaching pupils to solve verbal problems in arithmetic are feasible, but the effectiveness of these methods in practice depends to a large extent upon the zeal and skill of the teachers using them.

CHAPTER V

METHODS OF DIAGNOSIS AND REMEDIAL TREATMENT

"Diagnosis" is the term used to designate the methods by which specific disabilities of pupils are discovered. "Remedial treatment" designates the methods used in eliminating these specific disabilities. In the experiments on diagnosis and remedial treatment in arithmetic attempts have been made to determine the effectiveness of a variety of methods of diagnosis and of a variety of methods of remedial treatment. The experimental factor in these experiments may be characterized as exceedingly complex. Usually the factor includes a somewhat complicated procedure of diagnosis, still more complicated procedures of remedial treatment, and aspects more properly designated as "motivation devices." In none of the experiments does the experimental factor approach the specificity essential in order to give definite meaning to the findings.

Summary of reported conclusions. That diagnosis and remedial instruction are effective procedures in arithmetic is indicated in the investigations of Merton, and others.¹ Kallom,² Morton,³ Smith,⁴ Stevenson,⁵ Yeager,⁶ Buswell and John,⁷ Sister Kathleen,⁸ O'Brien,⁹ Otto,¹⁰ Clemens and Neubauer,¹¹ Neal and Foster,¹² Brownell,¹³ Chase,¹⁴ Gabbert,¹⁵ Guiler,¹⁶ Lazar,¹⁷ Soth,¹⁸ and Stone.¹⁹ It does not

¹Merton, E. L., et al. "Remedial Work in Arithmetic," *Second Yearbook of the Department of Elementary School Principals*. Washington: National Education Association, 1923, p. 395-429. (74)

²Kallom, A. W. "The Importance of Diagnosis in Educational Measurement," *Journal of Educational Psychology*, 10:1-12, January, 1919. (53)

³Morton, R. L. "An Analysis of Pupils' Errors in Fractions," *Journal of Educational Research*, 9:117-25, February, 1924. (81)

⁴Smith, J. H. "Individual Variations in Arithmetic," *Elementary School Journal*, 17:195-200, November, 1916. (107)

⁵Stevenson, P. R. "Increasing the Ability of Pupils to Solve Arithmetic Problems," *Educational Research Bulletin* (Ohio State University), 3:267-70, October, 1924. (112)

⁶Yeager, William. "Diagnosis and Remedial Measures in Arithmetic," *Twelfth Annual Schoolmen's Week Proceedings*. Philadelphia: University of Pennsylvania, 1925, p. 296-305. (127)

⁷Buswell, G. T. and John, Lenore. "Diagnostic Studies in Arithmetic," *Supplementary Educational Monographs*, No. 30. Chicago: University of Chicago Press, 1926. 212 p. (20)

⁸Kathleen, Sister M. "Some Results of Remedial Instruction Following the Use of Diagnostic Arithmetic Tests," *Catholic Educational Review*, 24:19-27, January, 1926. (54)

⁹O'Brien, F. P. "Improvement of Instruction in Arithmetic," *University of Kansas Bulletin of Education*, Vol. 1, No. 4. Lawrence: University of Kansas, 1927. 42 p. (93)

¹⁰Otto, H. J. "Remedial Instruction in Arithmetic," *Elementary School Journal*, 28:124-33, October, 1927. Also in: *Journal of the National Education Association*, 17:87-89, March, 1928. (96)

¹¹Clemens, P. B. and Neubauer, P. F. "A Supervision Project in Multiplication," *Journal of Educational Research*, 18:387-96, December, 1928. (28)

¹²Neal, E. A. and Foster, Inez. "An Experiment with Remedial Work in Common Fractions," *Elementary School Journal*, 29:280-83, December, 1928. (87)

¹³Brownell, W. A. "Remedial Cases in Arithmetic," *Peabody Journal of Education*, 7:100-107, September, 1929. (14)

¹⁴Chase, V. E. "The Diagnosis and Treatment of Some Common Difficulties in Solving Arithmetic Problems," *Journal of Educational Research*, 20:335-42, December, 1929. (24)

¹⁵Gabbert, M. L. "Remedial Cases in Arithmetic; Case 1," *Peabody Journal of Education*, 7:147-55, November, 1929. (39)

¹⁶Guiler, W. S. "Improving Computational Ability," *Elementary School Journal*, 30:111-16, October, 1929. (43)

seem worth while to present in detail the reported conclusions of all of these investigations. The conclusions of the single-group experiments and case studies contribute to our understanding of the effectiveness of discovering the individual arithmetical disabilities of pupils by means of diagnostic tests and by means of first-hand observation of the work of the pupil in which he is requested to think aloud in performing the fundamental operations or in solving problems.²⁰ The conclusions of these investigations also contribute to our understanding of the effectiveness of intensive and zealous instruction to eliminate the disabilities so discovered, either through the use of practice materials prepared in advance or informally at the time. These conclusions, important as they are, do not contribute materially, however, to our knowledge with respect to the relative effectiveness of the various methods of diagnostic and remedial treatment.

The conclusions of the controlled experiments contribute, in some measure, to our knowledge of the relative effectiveness of the various methods of diagnostic and remedial treatment. Smith (107) reported that class drill, supplemented by individual assistance on points of weakness revealed by diagnostic tests, is more effective than class drill with extra drill periods provided for the slow pupils who were drilled in groups rather than individually and class drill in which explanations were made only with respect to the group as a whole. Sister Kathleen (54) reported that remedial treatment is more effective when based on analysis and classification of the errors made on the test than when based only on class medians on the test. Neal and Foster (87) have reported that "organized practice material in the hands of the children, with provision for the diagnosis of difficulties and remedial work, is more effective in economy of the teacher's time and of the children's time and in final results in maintaining skill in the manipulation of common fractions than is the usual practice provided by the teacher." The conclusion of Stone (113) that diagnostic and practice tests produce "greater gains in ability to reason in arithmetic than does the regular work in arithmetic that the tests may displace in classroom use" agrees with that of Neal and Foster (87).

¹⁷Lazar, May. *Diagnostic and Remedial Work in Arithmetic Fundamentals for Intermediate Grades*. New York: Board of Education, 1928. 204 p. (66)

¹⁸Soth, M. R. "A Study of a Pupil Retarded in Arithmetic," *Elementary School Journal*, 29:439-42, February, 1929. (108)

¹⁹Stone, C. W. "An Experimental Study in Improving Ability to Reason in Arithmetic," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 589-99. (113)

Osburn and Foltz have recently reported an investigation in which a large amount of permanent improvement resulted from a program of inventory tests and special individual drills. See:

Osburn, W. J. and Foltz, P. J. "Permanence of Improvement in the Fundamentals of Arithmetic," *Educational Research Bulletin* (Ohio State University), 10:227-34, April 29, 1931.

²⁰The latter method of diagnosis is especially advocated by Buswell and John (20).

Evaluation of the investigations. The studies reported by Merton, and others, (74) and by Yeager (127) are to be characterized as "descriptive accounts of what is going on in some school." Some quantitative data are given and some comparisons in achievement of different classes are reported, but it is not possible to justify the labeling of such investigations "experiments." The studies of Kallom (53), Brownell (14), Chase (24), Gabbert (39), and Soth (108) were based on data secured from the following numbers of cases: 3, 4, 17, 1, 1. Descriptive accounts of what is taking place in schools and reports of case studies are interesting. They should be very suggestive to teachers in practice. It is impossible, however, to generalize from data so restricted.

Morton (81), Stevenson (112), O'Brien (93), Otto (96), Clemens and Neubauer (28), Guiler (43), and Lazar (66) conducted single-group experiments. Morton (81) used one group of thirty-six eighth-grade pupils for a period of five months. He measured the improvement of these pupils as a result of diagnostic and remedial treatment by means of tests constructed by himself. The substantial gains shown may not with certainty be ascribed to the experimental factor, because of the failure to employ a control group. The single-group experiment of Stevenson (112) was described and evaluated, rather unfavorably, in the previous chapter.²¹

O'Brien (93) used 357 pupils in the seventh, eighth, ninth, and tenth grades of three small school systems. After an initial program of mental and achievement testing, diagnosis was made with respect to "mental ability, previous schooling, achievement in various phases of the subject, and specific types of errors or difficulties which characterized the students' work." The program of remedial instruction was based on the weaknesses discovered by the tests. Pupils were informed of their individual weaknesses, and the teachers were provided with general and detailed suggestions for carrying out the remedial instruction. They were also provided with advice in conferences and with information in the form of abstracts of selected articles in current literature. At the end of five months the final tests were administered. While the increases in achievement are large, it is difficult to ascribe these increases to any specific experimental factor. No control groups were used, and it is evident that the pupils were subjected to a complex of factors.

Otto (96) used a single group of nine fourth-grade pupils for a period of seven months. Achievement was measured by diagnostic tests, and remedial treatment was provided by means of prepared

²¹See pages 58 to 67.

practice materials, but, again, because of lack of control, it is impossible to say how much of the improvement found is to be ascribed to the experimental factor. Clemens and Neubauer (28) employed a single group of 425 fourth-, fifth-, sixth-, seventh-, and eighth-grade pupils in twelve elementary schools of one city. Tests were constructed by the authors which covered forty-two multiplication difficulties. Tests were administered four times: (1) at the beginning of the experiment, (2) at the end of a week, (3) at the end of two more weeks, and (4) at the end of three months from the administration of the third test. "Individual help was given to each pupil who failed to obtain a perfect score in the first test. After correcting the child's error and showing him how to work the example correctly, the teacher gave him the drill card designed to meet his difficulty." Substantial gains in achievement were indicated by the test results, but failure to use a control group again makes it impossible to determine how much of this gain is to be ascribed to the experimental factor. Guiler (43) used a single group of ten seventh-grade pupils for one hour a week for twelve weeks. An analysis was made of the errors of these pupils on the diagnostic tests used, and remedial instruction adapted to individual needs was provided. The gains in achievement were measured by several standardized arithmetic tests, but it must be repeated again that failure to use a control group renders the conclusions of doubtful dependability.

Lazar (66) used a single group of forty-three sixth-grade pupils. The initial status of these pupils was determined by means of intelligence tests, of standardized arithmetic achievement tests, of a diagnostic arithmetic test, and by individual observation and oral examination. Ten minutes of the daily arithmetic period were devoted to remedial work characterized by the experimenter as follows: (1) Specific instruction on class or individual weaknesses as determined by diagnosis was given; (2) the Courtis Standard Practice Tests were used for drill on the operations in which deficiencies were shown; (3) supplementary material was devised to overcome difficulties with the addition combinations, with long division, and with fractions; (4) the pupils were taught how to make records and graphs to show their achievement, and the teacher made graphs of the class achievement; (5) training the pupils to have the proper attitude toward their deficiencies was an important phase of the work. At the end of five months the initial arithmetic tests were again administered. The gains in achievement appear to be "statistically" significant. While a control group was not used, some of the functions of a control group were attained by comparison of the experimental results with test

norms. The experiment is to be commended for its comprehensive and intensive nature, but Lazar's experiment deserves criticism similar to that applied to the experiment of O'Brien (93)—the experimental factor was exceedingly complex.

Buswell and John (20) investigated the problem of arithmetical diagnosis by means of two types of laboratory technique and by means of a comprehensive single-group experiment. In the laboratory study of eye-movements in column addition two fourth-grade, eight fifth-grade, and seven sixth-grade pupils were used. In addition to these groups of children, three adults were used. In the report of this research data are given in graphic form, which are dependable evidence with respect to the nature of eye-movements in column addition. This evidence emphasizes the need for diagnosis in arithmetical instruction.

The second laboratory investigation, in which thirty subjects were used, was conducted by means of dictaphone and kymograph apparatus. Time analyses were made of the four fundamental operations. What each child was asked to do is described in the following quotation:

The children who participated in the experiment were seated one at a time at a table on which was a sheet of paper. On this paper were typewritten the examples which they were to work. The only piece of apparatus in the room was a specially constructed telephone transmitter, which was clamped to the edge of the table. The experimenter sat beside the child and instructed him as to his procedure. The child was asked to give his partial answers aloud and also to say the digits which he wrote on the paper at the same time that he wrote them. In the case of an example in column addition the child was instructed to give each of the sums as he proceeded down the column.

The sound of the child's voice was reproduced by an amplifier in another room and recorded by means of a dictaphone. These records were then "transcribed on kymograph paper by using an electric time-marker and a telegraph key." The kymograph record may be described as follows. One line broken at regular intervals showed the time elapsed in intervals of fifths of a second. The second line, broken at irregular intervals revealed the time required for each partial answer. To illustrate by data secured from one child it was found that the child in adding a single column of thirteen digits required three-fifths of a second each to add $4 + 9$, $13 + 3$, and $16 + 2$. He required $19/5$ of a second to add the combination $29 + 3$. Data relative to time required to perform the fundamental operations for all the subjects are presented in tabular form in the monograph. An examination of the description of the techniques used gives no reason to doubt the reliability of these data. They are additional evidence of the need for diagnosis in arithmetical instruction.

Buswell and John used a single-group of 303 children, in nine classes, in twelve elementary schools. In a preliminary study they used a single group of 250 children in the third, fourth, fifth, and sixth grades. Diagnostic sheets, for each of the fundamental processes, were followed by remedial treatment administered by the teachers to suit the individual needs of the pupils. The Cleveland Survey Test was administered before and after the ten-weeks' period of diagnosis and remedial treatment, and substantial gains were found. Buswell and John hold that these gains may be ascribed to the experimental factor, even though a control group was not used. They state:

Owing to the lack of a refined technique in carrying on the experiment, a small difference between the actual improvement shown and the normal expected improvement cannot be considered significant. However, if the difference is fairly large, it seems fair to conclude that the difference is due to the diagnostic procedure and remedial instruction given by the teacher.

If this contention is accepted as correct, the evaluation of the dependability of the conclusions of the other single-group experiments must be modified. The gains in achievement were, without exception, large. The present writers do not feel, however, that the conclusions derived from data secured by single-group experimentation can be as satisfying, other things being equal, as those obtained from controlled experimentation. Obviously, it is impossible to determine how much of the gains in achievement was due to inherent qualities in the methods of diagnosis and remedial treatment and how much was due to additional and zealous instruction and to the mere drill afforded.

Control groups were used in the experiments of Smith (107), Sister Kathleen (54), Neal and Foster (87), and Stone (113). The experiment of Smith (107) has already been described and evaluated somewhat unfavorably.²² Sister Kathleen (54) used two groups of fifty sixth- and seventh-grade pupils in neighboring schools in her investigation of the relative effectiveness of remedial treatment based on analysis and classification of the errors made on a diagnostic test and remedial treatment based only on class medians on the test. She stated with respect to equivalence that the groups were "about the same average mental ability." The differences in gains which are not highly "statistically" significant were measured by the Woody-McCall Mixed Fundamentals Test, Forms I and II. The conclusions of Sister Kathleen seem to be somewhat more dependable than those of Smith (107), but the techniques used in this experiment were not without criticism. There is evidence of failure to control important non-experimental factors, particularly the factor of zeal on the part of the teachers.

²²See pages 30 to 33.

Neal and Foster (87) used approximately six hundred experimental and approximately four hundred control pupils in the fifth grade. These groups were not equivalent according to the initial-test scores, but allowance for non-equivalence is made in interpreting the results. The pupils in the larger group used "organized practice material, with provision for diagnostic and remedial work," while the pupils in the smaller group had "the usual practice provided by the teacher." The experiment lasted three months. The differences in gains in achievement, which are possibly "statistically" significant, were measured by the Stanford Achievement Test, Forms A and B, and by an informal fraction test prepared by the investigators. The experimentation deserves commendation with respect to the directions given participating teachers by means of mimeographed sheets. The conclusions stated would be more satisfying to the critical reader if appropriate restrictions had been made in addition to the recognition given to faulty equivalence.

Stone (113) made comparisons between groups of paired fifth-, sixth-, seventh-, and eighth-grade pupils of various sizes. In his preliminary trial 175 pairs of equivalent pupils were used. In his main trial comparisons were made between a total of 172 pairs. Other comparisons were made without resorting to pairing. The pupils participating in the experiment were located in twenty-three schools of five school systems. These pupils were paired with respect to arithmetic scores, mental age, chronological age, and school grade. Pairs were located in the same school systems. The pupils in the experimental groups had the benefit of a program of diagnostic and practice tests described by the experimenter as follows:

The *diagnostic tests* were designed to accompany the survey tests. Their purpose is to afford more precise means of locating each pupil's difficulties in arithmetical reasoning. They enable each pupil to think, by graduated steps, into and through his individual difficulty. The *practice tests* were designed to follow the diagnostic tests. Their purpose is to afford needed practice on specific difficulties, as located by survey and diagnostic tests. They enable each pupil to rethink the reasoning involved in his individual difficulty.

The pupils of the control group had the regular work in arithmetic without the benefit of a program of diagnosis and remedial treatment. The experiment lasted for five weeks. Gains in achievement were measured by the Stone Survey Tests I and II and by the Stone Reasoning Tests in Arithmetic. The differences in gains appear to be "statistically" significant. The chief criticism with respect to this experiment concerns the validity of the measuring instruments used. It seems possible that the tests may have been more valid with respect to the abilities engendered by the practice material. If this

was the case, some of the differences in gains should be attributed to this cause. The techniques used in this experiment are for the most part very commendable, especially those used in securing a representative sample and equivalent groups. The conclusions in favor of the diagnostic and remedial methods used with the experimental pupils are stated conservatively and as such seem quite dependable.

Justified conclusions. The generalization seems justified that diagnosis and remedial treatment should be recognized as necessary phases of instruction in arithmetic. The conclusions relative to the methods of diagnosis and remedial instruction are less certain. It seems evident from the comprehensive investigation of Buswell and John (20) that individual diagnosis and remedial instruction adapted to the needs of individual pupils are most effective. Other investigators obtained good results by means of diagnostic tests and practice material placed in the hands of the pupils, with less individual attention being given. There seems to be no reason to doubt that such methods are effective. Further research is needed, however, before it may be said that such methods are as effective as, or more effective than, methods in which emphasis is placed on direct observation of the pupil engaged in arithmetical learning activity and in which immediate provision of remedial instruction for the disabilities is discovered. It is quite evident that more attention should be given, in experimental evaluations of diagnostic and remedial methods, to the evaluation of specific aspects of such instruction rather than to evaluation of a complex of factors.

CHAPTER VI

METHODS OF TEACHING READING OF ARITHMETICAL SUBJECT-MATTER

It is fairly well known that children differ in their abilities to read various types of subject-matter. The reading of examples and of verbal problems in arithmetic involves the use of abilities quite different from those used in reading historical description or exposition. The research referred to in the first part of this chapter indicates the necessity of recognizing the significance of unique reading skills as factors in arithmetical achievement. The small number of experimental evaluations of methods of teaching the reading of arithmetical subject-matter is an indication that this problem has not received wide recognition among research workers in the field of arithmetic. One of the experiments described deals with the effectiveness of general training in reading. The second experiment deals with the effectiveness of a questioning method. It is also an attempt to evaluate dramatization and story telling as means of teaching the reading of verbal problems. In the third experiment, instructions in reading were included on the problem solution sheets provided for the pupils. There is need for an evaluation of a method which is more likely to engender the specific reading abilities required for arithmetical subject-matter.

Summary of reported conclusions. The necessity of instructing pupils in the reading of arithmetical subject-matter has been shown in a number of studies. Buswell and John,¹ Brooks,² Chase,³ Edaño,⁴ and Partridge⁵ have reported that a technical vocabulary is needed by children engaged in arithmetical learning activity. The conclusion stated by Chase (23) is typical:

... the investigation here recorded has shown after careful study of numerous textbooks, that many problems involve conditions that are quite untrue to life; that many of the words used are quite unknown to the one hundred children tested; and finally that forty-five experienced teachers from various school systems have found the subject-matter and vocabularies of the various texts which they have used quite unsuited to the capacities of their pupils.

¹Buswell, G. T. and John, Lenore. "The Vocabulary of Arithmetic," *Supplementary Educational Monographs*, No. 38. Chicago: University of Chicago Press, 1931. 146 p. (21)

²Brooks, S. S. "A Study of the Technical and Semi-Technical Vocabulary of Arithmetic," *Educational Research Bulletin* (Ohio State University), 5:219-22, May 26, 1926. (12)

³Chase, S. E. "Waste in Arithmetic," *Teachers College Record*, 18:360-70, September, 1917. (23)

⁴Edaño, Tiburcio. "An Analysis of Arithmetic Textual Matter," *Philippine Public Schools*, 1:81-84, February, 1928. (36)

⁵Partridge, C. M. "Number Needs in Children's Reading Activities," *Elementary School Journal*, 26:356-66, January, 1926. (99)

Several studies of errors made by pupils in the solution of arithmetical problems indicate that reading disability is an important cause of errors.⁶ Studies of the correlation between arithmetical ability and reading ability seem to indicate that a small but "statistically" significant correlation exists.⁷ In certain discussions of measurement in arithmetic it has been indicated that arithmetical achievement is in part a function of reading ability.⁸ In the opinion of the present writers the most significant evidence relative to the importance of instructing pupils to read arithmetic is to be found in the laboratory studies of Buswell and John⁹ and of Terry.¹⁰ The latter investigator has stated some suggestions for instructing pupils in reading arithmetical problems which seem worthy of quotation:

1. Pupils should be taught to distinguish between the first reading and the re-reading phases in their attack on problems.
2. They should learn to consider numerals and the accompanying descriptive conditions as different elements of a problem and separable for reading purposes.
3. During the first reading, they should devote their attention to the conditions of the problem.
4. At the same time skill should be developed in partial reading of numerals.
5. While this skill is being acquired, pupils should be apprised of the essential similarity between the conditions of the problem and such details of the numerals as are perceived by partial reading.¹¹

Experimental investigations of methods of instructing pupils to read arithmetical subject-matter have been reported by Newcomb¹²

⁶Hydle, L. L. and Clapp, F. L. "Elements of Difficulty in the Interpretation of Concrete Problems in Arithmetic," *Bureau of Educational Research Bulletin* No. 9. Madison: University of Wisconsin, 1927. 84 p. (50)

John, Lenore. "Difficulties in Solving Problems in Arithmetic," *Elementary School Journal*, 31:202-15, November, 1930. (51)

Morton, R. L. "An Analysis of Errors in the Solution of Arithmetic Problems," *Educational Research Bulletin* (Ohio State University), 4:187-90, April 29, 1925. (82)

Stevenson, P. R. "Increasing the Ability of Pupils to Solve Arithmetic Problems," *Educational Research Bulletin* (Ohio State University), 3:267-70, October 15, 1924. (112)

Stevenson, P. R. "Difficulties in Problem Solving," *Journal of Educational Research*, 11:95-103, February, 1925. (111)

⁷Hackler, J. M. "The Relation between Successful Progress in Mathematics and the Ability to Read and Understand, and the Factors that Contribute to Success or Failure in Mathematics," Unpublished master's thesis in Education. Chicago: University of Chicago, 1921. 82 p. (44)

Harlan, C. L. "Years in School and Achievements in Reading and Arithmetic," *Journal of Educational Research*, 8:145-49, September, 1923. (48)

Wheat, H. G. "The Relative Merits of Conventional and Imaginative Types of Problems in Arithmetic," *Teachers College, Columbia University Contributions to Education*, No. 359. New York: Bureau of Publications, Teachers College, Columbia University, 1929. 124 p. (121)

⁸Dawson, C. D. "Some Results in Using Starch's Arithmetic Reasoning Test," *Journal of Educational Research*, 2:677-78, October, 1920. (34)

Monroe, W. S. "The Derivation of Reasoning Tests in Arithmetic," *School and Society*, 8:295-99, 324-29; September 7, 14, 1918. (78)

⁹Buswell, G. T. and John, Lenore. "Diagnostic Studies in Arithmetic," *Supplementary Educational Monographs*, No. 30. Chicago: University of Chicago Press, 1926. 212 p. (20)

¹⁰Terry, P. W. "How Numerals Are Read: An Experimental Study of the Reading of Isolated Numerals in Arithmetical Problems," *Supplementary Educational Monographs*, No. 18. Chicago: University of Chicago Press, 1922. 110 p. (115) See also:

Terry, P. W. "The Reading Problem in Arithmetic," *Journal of Educational Psychology*, 12:365-77, October, 1921. (A Summary of the monograph referred to above.) (116)

¹¹Terry, P. W. "How Numerals Are Read: An Experimental Study of the Reading of Isolated Numerals in Arithmetical Problems," *Supplementary Educational Monographs*, No. 18. Chicago: University of Chicago Press, 1922, p. 101. (115)

¹²Newcomb, R. S. "Teaching Pupils How to Solve Problems in Arithmetic," *Elementary School Journal*, 23:183-89, November, 1922. (90)

Wilson,¹³ and Lessenger.¹⁴ The pupils in the experiment of Newcomb (90) were given instructions in reading problems on problem solution sheets, while in the experiment of Wilson (122) the pupils were taught to read problems by a questioning method and through dramatization and story telling. Lessenger (67) reported an experiment where general reading instruction was the experimental factor. These experiments lead to the general conclusion that reading instruction increases significantly the ability of pupils to solve arithmetical problems.

2. Evaluation of experiments. The investigations of Brooks (12), Chase (23), Edaño (36), and Partridge (99) were analytical rather than experimental in character. The need for instruction in reading was inferred from analyses of arithmetical materials of instruction. The investigations of Hydle and Clapp (50), John (51), Morton (82), Stevenson (111), and Stevenson (112) were also analytical in nature, but the analysis was made of pupil responses to arithmetical problems. Buswell and John (21) prepared group tests of arithmetical vocabulary and administered them to 1500 fourth-, fifth-, and sixth-grade pupils in several school systems. Their findings are probably the most significant in this group.

It is evident that the analytical investigations are limited by the inferences which had to be made. One may not be sure from observing a mistake made in a problem whether the cause of the faulty response was lack of reading ability or lack of some other ability. For example, the written performances of two pupils may be identical and thus not indicative of the fact that one of the pupils was handicapped by arithmetic disability while the other failed to solve the problem correctly because of reading disability.

Hackler (44) and Wheat (121) indicated the importance of reading ability in arithmetic learning activity by typical correlation techniques, and Harlan (48) showed that arithmetic and reading ability tend to occur together, indicating his correlation in graphic form. The correlation studies of Hackler (44), Wheat (121),¹⁵ and Harlan (48) are limited in dependability in the sense that all correlation studies are limited when the attempt is made to interpret them in terms of cause and effect. The raw coefficients obtained between arithmetic scores and reading scores are probably due in a large measure to the common factor of intelligence. If an attempt is made to partial out intelligence, the coefficient so obtained may be too much

¹³Wilson, Estaline. "Improving the Ability to Read Arithmetic Problems," *Elementary School Journal*, 22:380-86, January, 1922. (122)

¹⁴Lessenger, W. E. "Reading Difficulties in Arithmetical Computation," *Journal of Educational Research*, 11:287-91, April, 1925. (67)

¹⁵See page 57 for unfavorable criticism of Wheat's use of correlation methods.

reduced. Intelligence as represented in the intelligence score usually obtained includes reading ability. Partial correlation would not separate the two effectively, and the partial coefficient would of necessity be low.¹⁶

The laboratory investigation of Buswell and John (20) has already been described and favorably evaluated.¹⁷ Terry (115) used similar techniques. A portion of his data was secured by having his subjects record by means of a telegraph key and kymograph apparatus the time spent in the first reading and in the re-reading of arithmetical problems. The following data secured from one subject on one problem are illustrative:

- 7.6 seconds—time required for first reading
- 1.4 seconds—time required to re-read one numeral
- 2.4 seconds—time required to re-read another numeral
- .2 seconds—time required to re-read last sentence

Additional data were secured by means of eye-movement apparatus. All of Terry's data appear reliable evidence of the important function of reading ability in solving arithmetical problems. The suggestions made by Terry with respect to instruction in reading arithmetical problems may be regarded, however, only as suggestions. Terry has not shown by experimental trial that the method suggested is effective in increasing reading ability with respect to arithmetical problems.

The experiment of Newcomb (90) has already been described and criticized with respect to lack of representativeness of pupils used, lack of equivalence, and failure to secure adequate control of non-experimental factors.¹⁸ Wilson (122) used one group of thirty-four sixth-grade pupils of relatively low intelligence. These pupils were given the Stone Reasoning Test at the beginning of the experiment and were taught to read problems by a questioning method for twelve minutes three times a week for five weeks; at the end of this time they were tested again. The significant increase in achievement may not be ascribed with certainty to the experimental factor. Wilson reported similar results for instruction by which the children were directed to convert problems into stories and to dramatize them. One wonders how much of the reading ability so engendered would transfer to ordinary problem-solving activity.

Lessenger (67) used data collected from a single group of 111

¹⁶For a discussion of the limitations of correlation methods in causal investigations, see: Burks, B. S. "On the Inadequacy of the Partial and Multiple Correlation Technique," *Journal of Educational Psychology*, 17:532-40, 625-30; November, December, 1926. See also: Dunlap, J. W. and Cureton, E. E. "On the Analysis of Causation," *Journal of Educational Psychology*, 21:657-79, December, 1930.

¹⁷See pages 72 and 73.

¹⁸See pages 61 and 62.

pupils in Grades III to VIII, inclusive. Analysis of the arithmetical computation scores on the first test administered to the pupils showed a mean loss in arithmetical age of 6.1 months because of faulty reading. After a year of intensive general training in reading, analysis of the final-test results showed a mean loss in arithmetic age due to faulty reading of only .7 months. Classification and tabulation of the data secured from the initially good and poor readers revealed a superior gain in arithmetical age for the poorer readers. The investigator attributes this superior gain to the general training in reading. It is evident that this study is to be characterized as a rather crude experiment. No control group was used, and for this reason it is difficult to ascribe the improvement noted to the experimental factor.

Justified conclusions. The conclusion seems justified that reading ability is an important factor in arithmetical achievement. The magnitude of its influence in arithmetical achievement is not known, but the investigations of Buswell and John (20) and of Terry (115) indicate that it is a very important influence. This being the case, it seems justifiable to say that pupils should receive instruction in reading arithmetical subject-matter. Further research must be conducted, however, before a dependable conclusion may be stated relative to the nature of the most effective instruction.

CHAPTER VII

MOTIVATION OF LEARNING IN ARITHMETIC

The assignment of learning exercises which are of immediate interest to pupils is recognized as a basic procedure in securing motivation of learning activity in the various school subjects. Attention is given, therefore, in this chapter to research on the stimulating effect of various types of learning exercises in arithmetic. Certain supplementary procedures for securing intensive effort and persistence in learning have been shown to be effective in the general research on motivation.¹ Some of these procedures have been employed as experimental factors of experiments in the field of arithmetic. These supplementary procedures are definite goals or objectives, knowledge of status or progress, competition, commendation, and reproof.

Summary of reported conclusions. The conclusions of investigations relating to the motivation of learning in arithmetic are summarized under the following heads: (1) effect of types of learning exercises, (2) effect of definite goals, (3) effect of knowledge of status or progress, (4) effect of competition, (5) effect of commendation and reproof. As will be noted, several of the investigations involved more than one motivation procedure. Consequently such studies will appear under two or more heads.

Number games,² problems presented in story form,³ dramatization of activities that create arithmetical problems,⁴ problems relating to the out-of-school life of pupils,⁵ and problems which the pupils believe they can solve successfully⁶ have been reported as effective in stimulating learning activity.

The stimulating effect of definite goals is usually involved in the use of standardized tests, especially when the attention of the pupils

¹For a summary of this research, see:

Monroe, W. S. and Engelhart, M. D. "Stimulating Learning Activity," *University of Illinois Bulletin*, Vol. 28, No. 1, Bureau of Educational Research Bulletin No. 51. Urbana: University of Illinois, 1930, p. 42-58.

²Steinway, L. S. "An Experiment in Games Involving a Knowledge of Number," *Teachers College Record*, 19:43-53, January, 1918. (110)

³Wilson, Estaline. "Improving the Ability to Read Arithmetic Problems," *Elementary School Journal*, 22:380-86, January, 1922. (122)

⁴*Loc. cit.*

Reavis, W. C. "The Social Motive in the Teaching of Arithmetic," *Elementary School Journal*, 18:264-67, December, 1917. (101)

⁵Graham, V. O. "Arithmetic Reasoning Project and the Measurement of Improvement," *Chicago Principals' Club, Second Yearbook*. Chicago: Chicago Principals' Club, 1927, p. 86-87. (41)

Kulp, C. L. "A Method of Securing Real-Life Problems in the Fundamentals of Arithmetic," *Elementary School Journal*, 29:428-30, February, 1929. (64)

⁶Bowman, H. L. "The Relation of Reported Preference to Performance in Problem Solving," *University of Missouri Bulletin*, Vol. 30, No. 36, Education Series, No. 29. Columbia: University of Missouri, 1929. 52 p. (10)

is directed to the norms specified by the tests. Motivation of learning activity by means of administering standardized arithmetic tests has been reported by Ballou,⁷ Courtis,⁸ Krause,⁹ O'Brien,¹⁰ and Wertheimer.¹¹ In such cases it is likely that the attainment of a high score was recognized by the pupils as a definite goal.

In several investigations it is difficult to separate the effect of definite goals from the effect of knowledge of progress. The latter factor, however, has been reported as having a beneficial influence in arithmetical learning activity by Sheerin,¹² Richardson,¹³ Anthony and others,¹⁴ Chapman and Feder,¹⁵ Hahn and Thorndike,¹⁶ Kirby,¹⁷ and Panlasigui and Knight.¹⁸

The ease with which arithmetical achievement, especially in the field of calculation, is measured facilitates competition between individual pupils and between groups. Maller¹⁹ has reported that individual competition is the more effective. Hahn and Thorndike (46) have reported that directing each pupil to compete with his own record was found to be an effective motivating device in learning addition.

The motivating effects of commendation and reproof have been studied by Hurlock²⁰ and by Newcomb.²¹ The former found that although, in general, commendation is superior to reproof as a motivating procedure, girls are more affected by praise than boys, while boys are more affected by reproof than girls. She found also that older and younger children are about equal in responsiveness to praise and reproof, and that inferior children are most responsive to

⁷Ballou, F. W. "Improving Instruction through Educational Measurement," *Educational Administration and Supervision*, 2:354-67, June, 1916. (7)

⁸Courtis, S. A. "The Courtis Standard Tests in Boston, 1912-15: An Appraisal," *School Document* No. 15, 1916. Department of Educational Investigation and Measurement Bulletin No. 10, Boston: Boston Printing Department, 1916. 48 p. (33)

⁹Krause, A. K. "Why Monroe Diagnostic Tests in Arithmetic?" *Contributions to Education*, Vol. 2. Yonkers, New York: World Book Company, 1928, p. 15-17. (63)

¹⁰O'Brien, F. P. "Co-operative Experiment Pertaining to Instruction in Arithmetic," *American Education*, 31:219-21, February, 1928. (92)

¹¹Wertheimer, J. E. "Some Results of Monroe's Diagnostic Tests in Arithmetic," *Journal of Educational Psychology*, 11:109-12, February, 1920. (120)

¹²Sheerin, E. M. "Application of the Dalton Plan to Teaching Arithmetic," *Contributions to Education*, Vol. 2. Yonkers, New York: World Book Company, 1928, p. 18-22. (106)

¹³Richardson, J. W. "The Campaign Method in Elementary Education," *Journal of Educational Research*, 2:481-92, June, 1920. (104)

¹⁴Anthony, Kate, et al. "The Development of Proper Attitudes Towards School Work," *School and Society*, 2:926-34, December 25, 1915. (3)

¹⁵Chapman, J. C. and Feder, R. B. "The Effect of External Incentives on Improvement," *Journal of Educational Psychology*, 8:469-74, October, 1917. (22)

¹⁶Hahn, H. H. and Thorndike, E. L. "Some Results of Practice in Addition under School Conditions," *Journal of Educational Psychology*, 5:65-84, February, 1914. (46)

¹⁷Kirby, T. J. "Practice in the Case of School Children," *Teachers College, Columbia University Contributions to Education*, No. 58. New York: Bureau of Publications, Teachers College, Columbia University, 1913. 98 p. (57)

¹⁸Panlasigui, Isidoro, and Knight, F. B. "The Effect of Awareness of Success or Failure," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 611-19. (98)

¹⁹Maller, J. B. "Cooperation and Competition, An Experimental Study in Motivation," *Teachers College, Columbia University Contributions to Education*, No. 384. New York: Bureau of Publications, Teachers College, Columbia University, 1929. 176 p. (70)

²⁰Hurlock, E. B. "An Evaluation of Certain Incentives Used in School Work," *Journal of Educational Psychology*, 16:145-59, March, 1925. (49)

²¹Newcomb, R. S. "Securing the Maximum Amount of Work from Every Pupil," *Elementary School Journal*, 25:376-79, January, 1925. (89)

praise while superior children are most responsive to reproof. Newcomb (89) urged pupils to solve supplementary problems, and by commending them when they did so, secured effective results. In addition to informing pupils of their progress and of their goals, Kirby (57) secured motivation by commending the attainment of high scores.

Evaluation of experiments. Several of the investigations may be termed "uncontrolled" experiments. Graham (41) based his conclusions on the apparently successful results secured in his school system when the method of relating problems to the out-of-school life of the pupils advocated by him was tried out. He reports little quantitative data. Kulp (64) reports no quantitative data at all but describes the success in his school when a similar method was tried. Wilson (122) did not secure equivalence for the groups used in her experiment, and, hence, the relative merits of the methods of dramatization and story telling in connection with teaching verbal problems cannot be determined. She presents somewhat more quantitative data than Graham (41) in favor of the effectiveness of her methods. Steinway (110) used two groups of children in her investigation, reporting the effectiveness of securing motivation by number games, but, here again, the lack of equivalence and failure to use suitable measuring instruments makes it impossible to list this as other than a crude experiment.

The studies of Reavis (101), Sheerin (106), Richardson (104), and Newcomb (89) were single-group experiments. Reavis (101) used a single group of twenty-one eighth-grade pupils, organized the class as a bank in which such learning activities as exercises with stocks, bonds, deposits, and checks were engaged in, and measured the gain in achievement by means of an informal problem test administered at the close of the experimental instruction and again some months later. Sheerin (106) used a single group of unreported size for a period of four months. One aspect of the experimental factor was that of informing pupils of progress. No mention is made of any attempt to measure quantitatively the improvement ascribed to the method by the investigator. Richardson (104) used single groups of indefinitely reported size. In the first "campaign" ten intermediate-grade classes participated for a period of nine weeks. In the second campaign pupils in the fourth, fifth, sixth, seventh, and eighth grades of "some fifteen" schools participated. It is stated that the numbers in each grade ranged from 250 in the fourth grade to 150 in the eighth. This campaign lasted for six weeks. In the third campaign the fourth, fifth, and sixth grades took part, while the seventh and eighth served

in some measure as control groups. Newcomb (89) used a single group of seventh- and eighth-grade pupils of unreported size. The Courtis and Stone tests were administered before and after the experimental period. Substantial gains in achievement were reported, but because of the lack of control it is impossible to ascribe these gains with certainty to the experimental factor.

It should be apparent that all of these single-group experiments are open to serious criticism. One cannot determine to what extent the improvement was the result of the application of the experimental factor, since many other factors were operating. In Richardson's investigation (104) teacher zeal probably was an influential factor. Other criticisms may be mentioned. In most of the experiments the improvement was inadequately measured, if it was measured at all. For the most part these experiments may be characterized as "descriptive accounts of what is going on" in a certain school.²²

Hahn and Thorndike (46), Kirby (57), Chapman and Feder (22), Panlasigui and Knight (98), Hurlock (49), Maller (70), and Bowman (10) conducted controlled experiments. Those of Hahn and Thorndike (46) and of Kirby (57) were quite favorably evaluated in the section on the effect of distributing practice in drill on the fundamentals.²³

Chapman and Feder (22) used two groups of sixteen fifth-grade pupils. These groups were exercised ten minutes a day on an addition test, one minute a day on a cancellation test, and five minutes a day on a substitution test. One of the groups was subjected to such motivating influences as the following:

- (1) Each individual's results of the previous day were published.
- (2) On sheets presented for the day's work, the point reached on the last occasion by the subject was marked in heavy blue pencil.
- (3) The general improvement of the class was presented in the form of a graph.
- (4) Credits were given in the form of stars, It was understood that prizes of a merely nominal value were to be given at the end of the ten practice periods to the 50 per cent in Group A which had gained the greatest number of stars for efficiency and improvement.

Data secured for ten practice periods are presented in tabular and graphic form. The achievement of the motivated group in addition was certainly significantly superior to the achievement of the non-motivated group. The chief criticisms to be made of this experiment have to do with the complex experimental factor described above and the artificiality of conditions.

²²Judd would deny the label "educational research" to such writings. See: "What is Research," *School Review*, 34:488, September, 1926. (An editorial.)

²³See pages 45 to 48.

Panlasigui and Knight (98) used a total of 358 experimental pupils and an equal number of control pupils in the fourth grade of ten school systems in nine states. The pupils were paired with respect to arithmetic ability shown by the initial test. The degree to which equivalence was attained is indicated by the fact that the means, first and third quartiles, and standard deviations of the two distributions of initial-test scores were identical. The drill materials used by the experimental pupils differed from the materials used by the control pupils in that each pupil could determine his individual progress. Class progress charts were also provided for the pupils in the experimental group. With respect to the control of non-experimental factors the authors state that "serious attempts were made to minimize all unusual factors and to approximate normal conditions." It is unfortunate that the authors do not describe what these attempts were. The experiment continued for twenty weeks; at the end of this time the final test was administered. The difference in final-test means was 3.93 times its probable error and is an indication that the chances of the true difference having the same sign, or of being in the same direction, are approximately 286 to 1.²⁴ The data are interpreted also for the "top and bottom quarters" of all groups on the initial test, and other comparisons are made. It is evident that Panlasigui and Knight have reported an excellent experiment. What criticism may be made concerns such things as failure to report the reliability of the tests used and failure to state how non-experimental factors were controlled. It is possible that the conclusions are not sufficiently restricted with respect to limitations of the data, but until better evidence is reported to the contrary, it would seem that they may be accepted as dependable evidence of the effectiveness of stimulating arithmetical learning activity by insuring that pupils are aware of their progress.

Hurlock (49) used four groups of fourth- and sixth-grade pupils, two of twenty-six pupils each and two of twenty-seven pupils each. It is stated with respect to equivalence that "these groups were equal not only in initial ability as displayed on these tests in addition, but also in average age and number of boys and girls within each group." The first group was praised over a period of five days in the presence of other members of their classes. The second group was reproved under the same conditions, while the third group was ignored. It

²⁴286 to 1 are the chances when a difference is four times its probable error. Chances of at least 369 to 1 are usually accepted as "statistically" significant. See:

Monroe, W. S. and Engelhart, M. D. "Experimental Research in Education," *University of Illinois Bulletin*, Vol. 27, No. 32, Bureau of Educational Research Bulletin No. 48. Urbana: University of Illinois, 1930, p. 60, 66.

McCall, W. A. *How to Measure in Education*. New York: Macmillan Company, 1922, p. 404-05.

should be mentioned that the pupils of the third group heard the praise and reproof of the others. The fourth group was used as control and was tested in a separate room. Modifications of the addition test of the Courtis Research Tests in Arithmetic were administered each day for five days. The differences in test means are greatest when the praised group is compared to the control, the chances of significance in favor of praise being "10,000 in 10,000." When the reproofed group is compared with the control, the chances are "9,382 in 10,000" in favor of reproof, and when the ignored group is compared with the control, "5,338 in 10,000" in favor of hearing praise and reproof of others. The conclusions seem reasonably dependable from the standpoint of the conditions of the experiment. One wonders how significant they are for ordinary classroom practice. It is possible that praise and reproof are effective incentives to learning arithmetic in the typical class, but how effective they are must await experiments with less abnormal conditions.

Maller (70) used 814 experimental and 724 control pupils in Grades V to VIII. The experimental pupils, alternately stimulated by individual recognition and reward and by group or class recognition and reward, solved addition examples. The investigator states in this connection:

The tests of work for self and work for class were repeated twelve times, two minutes each. The motives of self and class were alternated six times, respectively. The problem of practice effect was thus practically eliminated. All conditions of work aside from the motives were identical.²⁵

The difference in favor of individual competition when compared with group competition was almost thirteen times its probable error. For the conditions of the experiment there is little reason to doubt the significance of this difference. The experimental conditions may be characterized as abnormal. It is doubtful whether competition would appeal to school children as a continuous diet in ordinary teaching. It is likely that its effectiveness would lessen with continued use.

The experiment of Bowman (10) has been described and favorably evaluated in the section on methods of teaching and learning verbal problems.²⁶

The data of the investigations of Ballou (7), Courtis (33), Krause (63), O'Brien (92), and Wertheimer (120) were secured by the administration of such standardized tests as those by Courtis and by Monroe. Increased achievement in arithmetic seem to result through repeated administration of such tests. The authors of the

²⁵Maller, *op. cit.*, p. 15.

²⁶See pages 52 to 58.

reports of these investigations ascribe some of the improvement to the stimulating effect of the tests. There is no means of showing, even by controlled experiments, which these investigations certainly were not, how much of the achievement may be ascribed to this factor. It is doubtful whether a controlled experiment could be set up which would satisfy the law of the single-variable, since it would be impossible to separate the motivating factor from the complex group of factors which systematic testing involves.

Anthony and others (3) secured their data from intensive case studies of three children. The studies which were conducted over a period of five months revealed the progress of the children by means of learning curves. Their conclusions in favor of the use of learning curves cannot be regarded as other than suggestive because of the small number of cases.

Justified conclusions. The only conclusion which may be offered as dependable is that knowledge of progress in arithmetical learning is an effective motivating influence. It does not seem to matter a great deal what methods the teacher uses to insure that pupils are aware of their success or failure. Individual learning curves, progress charts, test scores, and the like seem to be effective devices. The conclusions relative to commendation and reproof are less certain, but research in other subjects with respect to motivation seems to indicate that commendation is most effective, reproof somewhat effective, and both are more effective than no comment at all.²⁷

Evidence that certain devices and methods—namely, the project method, the Dalton plan, the use of games involving a knowledge of numbers, the telling of stories in connection with problems, the dramatization of the stories,²⁸ and the use of tests—are stimulating to learning activity in arithmetic is to be found in the single-group experiments. It should be noted that the evidence with respect to these methods and devices may not be regarded as highly dependable.

Formulation and presentation of appropriate learning exercises are possibly the most effective means of securing motivation of learning activity in arithmetic. It should not be inferred that pupil preference is the most important criterion in the devising of learning exercises.²⁹ It should be used as a criterion only after the test of compatibility with recognized objectives has been applied. The data of Bowman (10) reveal that belief in success causes preference. Capable instruc-

²⁷Monroe, W. S. and Engelhart, M. D. "Stimulating Learning Activity," *University of Illinois Bulletin*, Vol. 28, No. 1, Bureau of Educational Research Bulletin No. 51. Urbana: University of Illinois, 1930, p. 51-54.

²⁸The reader should compare this statement with the conclusions of Wheat (121). See pages 52 to 58.

²⁹The findings of Bowman (10) should be referred to in this connection. See pages 52 to 58.

tion by problems which are desirable from the standpoint of objectives should be effective in engendering preferences for such problems.

As a conclusion to this summary of the experiments on motivation in arithmetic the following statement taken from the monograph on motivation previously referred to seems pertinent:

If the teacher has a real interest in children and in teaching, if she approaches her pupils with the attitude that doing the exercises assigned is an interesting and challenging activity, the problem of motivation will tend to disappear. Motivation procedures and devices will be needed only to supplement the stimulating effect of other instructional procedures.³⁰

³⁰Monroe, and Engelhart, *op. cit.*, p. 58.

CHAPTER VIII

GENERAL SUMMARY AND CONCLUSION

This chapter begins with a list of the problems of the investigations summarized in the preceding chapters. The statement of each problem is followed by a note with respect to the reported conclusions. Where the statement is made that a reported conclusion is undependable, it may be inferred that the conclusion is unworthy of generalization. In some instances the note following the problem statement contains a remark relative to a possible, more correct solution of the problem. In the paragraphs following this list of problems an estimate is presented of the contribution of experimental research up to the present. This estimate is followed by suggestions for further research in this field and with a statement of the requirements for precise evaluation of instructional techniques in arithmetic. The chapter closes with a discussion of feasibility versus relative effectiveness of instructional techniques.

The problems studied. The following questions represent the problems of the arithmetic investigations summarized in the preceding chapters. While the questions are, for the most part, quite specific in character, it was felt that some synthesis was desirable. Where several investigations were made of practically the same problem, one problem statement was formulated to represent all of them. Where investigations were made of different aspects of the same problem, a compound statement was formulated to represent the aspects investigated.

1. What is the relative efficiency of upward versus downward addition?

The reported conclusion that the method of teaching pupils to add in the downward direction is superior in effectiveness to the method of teaching pupils to add in the upward direction seems undependable. It appears probable that there is no significant difference in effectiveness between the two methods.

2. What is the relative effectiveness of the following methods of teaching addition and subtraction: (1) Showing pupils how to perform the process with no consideration of generalization or of underlying principles; (2) helping pupils to formulate general methods of procedure from specific types taught and emphasizing these generalizations throughout the teaching; (3) teaching the reasons and principles underlying the specific types taught; (4) teaching both general methods and general principles?

The reported conclusion favoring (2) appears undependable. It seems reasonable that (4) should be superior in effectiveness to either (2) or (3).

3. What is the relative effectiveness of three minutes' instruction daily in generalizing groups of addition and subtraction combinations included within

twenty-minute practice periods in addition and subtraction and twenty-minute practice periods without the generalizing instruction?

Three minutes' daily instruction in generalizing groups of addition and subtraction combinations, within twenty-minute practice periods in addition and subtraction, is reported not to add significantly to the achievement engendered by the twenty-minute practice periods alone. The conclusion as stated is reasonably dependable, but it should not be inferred that generalizing instruction is inherently ineffective.

4. What is the effectiveness in performing the fundamental operations of "thinking results only?"

The reported conclusion that the method is effective is based on limited experimental evidence, but it seems reasonable that this method is effective since it tends toward the establishment of more direct mental processes.

5. What is the effectiveness of teaching pupils to break long columns into two parts and to add each part separately?

The reported conclusion that this method is effective is based on very limited experimental evidence. It seems reasonable that the method would engender undesirable addition habits.

6. What are the relative merits of adding digits in regular serial order and making mental combinations or rearrangements?

The reported conclusion favoring serial order is based on faulty experimental data. The conclusion, however, appears reasonable since excessive combination and rearrangement is likely to prove confusing to immature pupils.

7. What is the effectiveness of teaching pupils to check their answers in addition?

The reported conclusion favoring the effectiveness of this method is not supported by adequate experimental evidence. It appears reasonable, however, that checking is an effective means of securing accuracy in addition, and that the attainment of accuracy is worth the possible sacrifice in speed necessitated by checking.

8. What are the relative merits of the following methods of subtraction: (1) Subtractive or take-away in which borrowing or decomposition is used; (2) subtractive or take-away in which carrying or equal addition is used; (3) additive in which borrowing or decomposition is used; (4) additive in which carrying or equal addition is used.

The second of these four methods of teaching or learning subtraction is reported to be superior in effectiveness to the others. It seems reasonable to assume that all four of the methods are feasible and that there is no significant difference in their effectiveness.

9. What are the relative merits of the multiplicative method of division and the traditional method?

The multiplicative method is reported to be superior on the basis of inadequate experimental evidence. It seems reasonable to assume that the multiplicative method is not significantly more effective than the traditional method.

10. What is the effectiveness of using addition of fractions as a basis for teaching the multiplication of fractions?

The method is reported to be effective on the basis of faulty experimental data. It seems reasonable to assume, however, that the method is an effective one since it conforms to the principle of apperception.

11. What is the effectiveness of providing drill in the fundamental combinations as a means of increasing achievement in common and decimal fractions?

The evidence supporting the reported conclusion that the above method is effective is undependable. It appears reasonable, however, that the method is effective. It is self-evident that pupils are unlikely to have sufficient mastery of the four funda-

mentals that further drill will not increase their achievement where these skills are used.

12. What is the relative effectiveness of practice material so prepared that the type of percentage problem set for solution is apparent to the pupils and of the ordinary textbook material?

The reported conclusion favoring the prepared material is not supported by very dependable experimental evidence, but the conclusion appears reasonable. It is compatible with other findings respecting prepared practice material.

13. What is the effectiveness of teaching children to place the decimal point in a quotient by means of a general rule?

The conclusion which reports that the method is ineffective is based on scanty experimental evidence, but the relatively specific nature of the division abilities justifies the assumption that the conclusion is reasonably correct.

14. What, in learning division, is the relative effectiveness of the rules: "There are as many places in the quotient as those in the dividend exceed the divisor" and "First render the divisor an integer by multiplying both dividend and divisor by 10 or some power of 10. Then proceed as with integral divisors."

In learning division it is reported that use should be made of the rule, "First render the divisor an integer by multiplying both dividend and divisor by 10 or some power of 10, and then proceed as with integral divisors" rather than of the rule, "There are as many places in the quotient as those in the dividend exceed the divisor."

15. What is the effectiveness of the "method of unity" in teaching proportion?¹

The experimental evidence supporting the conclusion that the method is effective is not dependable. It seems reasonable to postulate, however, that the method is an effective one.

16. What is the relative effectiveness of memorizing tables of cubic and linear measure as compared with the effectiveness of using the facts of these tables in connection with problems?

It is reported that it is more effective for pupils to memorize tables of cubic and linear measure than to learn them through using the facts of these tables in connection with problems. It is a fairly well accepted principle of learning, however, that information learned through use is usually better retained than information learned in isolation from use.

17. What is the effect on achievement in arithmetical calculation of systematic drill in addition, subtraction, multiplication, and division?

The conclusion that systematic drill is effective is supported by comprehensive and reasonably dependable experimental evidence. It may be accepted as an established general principle.

18. What is the relative effectiveness of systematic versus incidental teaching of calculation?

The systematic method of teaching calculation is reported to be more effective than the incidental method. The incidental method of teaching calculation is also reported to be more effective than the systematic. The findings of research in other fields, and logical thinking would favor a combination of both methods, with possibly greater emphasis on the systematic.

19. What is the effectiveness of a combination of systematic and incidental methods of teaching calculation?

The conclusion that a combination of systematic and of incidental method of teaching calculation is effective is not based on highly dependable experimental

¹See page 26 for an illustration of this method.

evidence. It appears, however, to be a reasonably correct solution of the problem, since the incidental method should contribute motivation and the systematic method should insure the distribution of practice compatible with the recognized objectives of arithmetic.

20. What is the relative effectiveness of various types of drill materials which have been prepared by experts? How do these drill materials compare in effectiveness with those prepared informally by teachers?

The conclusions reported with respect to the relative effectiveness of the different prepared materials are undependable. The conclusions with respect to the superiority of the materials prepared by experts as compared with those prepared by teachers appear to be reasonably dependable.

21. What is the effectiveness of drill exercises in addition prepared in such a way that proportionate drill is given on the higher decades as compared with drill materials ordinarily used?

The conclusion favoring the prepared material is not supported by adequate experimental evidence. The conclusion seems, however, to be reasonably correct.

22. What is the effectiveness of teaching the one hundred multiplication combinations by means of text material alone with the teacher doing as little talking as possible?

The conclusion favoring the above method is not supported by sufficient experimental evidence. Further research is needed before it may be concluded that the teacher has no function in drill.

23. What is the relative effectiveness of drill material so constructed that practice is distributed over the number combinations and of drill material in which certain combinations are slighted?

The reported conclusion in favor of the material which provides distributed practice is supported by rather highly dependable experimental evidence.

24. What is the effectiveness of drill material so prepared that the amounts of practice provided on the number combinations are proportional to their difficulty?

The conclusion that drill material should be prepared in this way seems to be supported by reasonably acceptable experimental evidence.

25. Is it better to have pupils find mistakes among a group of examples of addition, multiplication, and subtraction combinations than to have them think only the correct associations?

The conclusion that it is better to have pupils think only the correct associations is not supported, in this instance, by dependable experimental evidence. It is a well accepted principle of learning, however, that it is more desirable for pupils to come in contact with that which will engender correct associations, than to come in contact with that which is likely to engender incorrect associations.

26. What is the effect on computational achievement of drill materials which provide practice in arithmetical reasoning?

The reported conclusion that such materials increase computational achievement, while not supported by adequate experimental evidence, appears, however, to be reasonably correct since it conforms to the Law of Exercise.

27. Should addition and subtraction be taught together or separately?

It is reported that addition and subtraction should be taught together. It is reasonable to assume that there should be separate teaching of addition and subtraction during the initial stages of learning, and mixed teaching for maintenance, or increase, of skill.

28. What is the relative effectiveness of drill material of mixed nature and drill material in which practice on addition, subtraction, multiplication, and division is provided for separately?

The conclusion that drill material of mixed nature is relatively more effective than drill material in which separate practice is provided for addition, subtraction, multiplication, and division is based on reasonably acceptable experimental evidence. It should be noted that such drill is reported to be effective for maintenance of skill and increase of skill. Its use is possibly unjustified before a certain level of attainment has been reached with each of the four fundamentals.

29. What is the optimum distribution of practice time in the fundamentals?

The reported conclusions with respect to this problem are not in close agreement, nor are they based on adequate experimental evidence. It would seem reasonable, however, to suggest that twenty-minute practice periods at daily intervals until mastery has been attained and shorter periods at longer intervals for the maintenance of skill, approximates the optimum distribution of practice time in the fundamentals. This suggestion would appear to be in agreement with findings of most of the research on distribution of practice time and with what is known concerning the curve of retention.

30. What are the relative effects of requests for speed and of requests for accuracy on achievement in the fundamentals?

The conclusion is reported that it is preferable to request accuracy of pupils rather than speed in the earlier stages of learning. After mastery has been attained speed may be requested. While this conclusion is not supported by acceptably dependable experimental evidence, it seems compatible with the principle that repetition of incorrect response should be avoided.

31. What are the characteristics of pupil responses to verbal problems in arithmetic? To what extent is the response the result of reflective or critical thinking?

The conclusion that pupil responses to verbal problems are usually characterized by lack of critical reflective thinking appears to be reasonably dependable.

32. What are the influences on problem-solving performance of the following characteristics of problem statements: familiar terminology, unfamiliar terminology, imaginative elements, irrelevant elements, size of numbers, amount of computation, and the like?

The conclusion that pupil responses to verbal problems are more satisfactory when they are stated in familiar terminology and without irrelevant elements appears reasonably dependable. The conclusion that responses are less likely to be satisfactory when problems are stated imaginatively is less dependable, but appears reasonable. The conclusions with respect to other aspects of problem statements are even less dependable. Further research is needed for determining what is most effective with respect to these aspects.

33. What is the effectiveness of providing pupils with systematic training in finding the facts pertaining to the problem, in deciding the processes to be used, and in finding the answer in round numbers?

Systematic training in finding the facts pertaining to the problem, in deciding the processes to be used, and in finding the answer in round numbers is reported to be an effective method of teaching verbal problems in arithmetic. The experimental evidence supporting this conclusion appears reasonably dependable.

34. What is the relative effectiveness of teaching pupils to solve problems by the graphic and by the conventional methods?

It is reported that it is more effective to teach pupils to solve verbal problems in arithmetic by the graphic than by the conventional methods. The experimental evidence supporting this conclusion is faulty, and it seems reasonable to assume that the graphic method would engender round-about and otherwise uneconomical habits.

35. What is the effectiveness of assigning large numbers of problems in teaching children to solve problems?

It is reported to be effective in increasing problem-solving achievement to assign

large numbers of problems. This conclusion, while not based on acceptable experimental data, agrees with the Law of Exercise.

36. What is the effectiveness of teaching pupils to see the analogies between difficult written problems and correspondingly easy oral problems?

The conclusion is reported that the method is not effective. Since this conclusion is not supported by very dependable experimental evidence, and since the method would appear to be compatible with the Law of Association, it would seem reasonable to suppose that the method is effective.

37. What is the value of diagnostic and remedial treatment in arithmetic?

Diagnostic and remedial treatment is highly effective in the field of arithmetic. The experimental evidence in support of this conclusion is comprehensive and reasonably dependable.

38. What is the relative effectiveness of individual diagnosis in which "first-hand observation is made of the actual work of the pupil" and diagnosis by means of diagnostic tests?

Conclusions have been reported in favor of both methods of diagnosis. Further research is needed to determine which method is relatively more effective. It seems reasonable that both methods are very feasible.

39. What is the relative effectiveness of remedial treatment in which pupils are given organized drill material affording practice of abilities diagnosed as weak and of informal material prepared by the teacher?

The conclusion favoring the expertly prepared remedial drill material is not supported by adequate experimental evidence, but it does conform with other conclusions respecting expertly prepared drill material.

40. To what extent is reading ability a factor in arithmetical achievement?

That reading is an important factor in arithmetic achievement seems reasonably well established. Further research is needed to show the precise magnitude of the influence of this factor.

41. What is the effectiveness of general training in reading in engendering greater achievement in arithmetic?

General training in reading is reported effective in engendering greater achievement in arithmetic. The experiment in which general training in reading constituted the experimental factor was very crude, but the conclusion is supported by the research which reveals that reading ability is a factor in arithmetical achievement.

42. What is the effectiveness of solution sheets containing information with respect to the manner of reading problems and containing spaces for recording of data useful at different stages in the solution of the problem?

Solution sheets containing information with respect to the manner of reading problems and containing spaces for the recording of data useful at different stages in the solution of problems are reported to be an effective device in teaching pupils to solve problems. While the experimental evidence is not of acceptable dependability, the method would seem to be feasible since more direction is given to the learning activity. It is possibly more desirable for the earlier rather than the later stages of learning to solve verbal problems.

43. What is the effectiveness of story-telling and dramatization in teaching pupils to read verbal problems in arithmetic?

Story-telling and dramatization are reported, on the basis of very limited experimental evidence, to be effective devices in teaching pupils to read verbal problems in arithmetic. This conclusion appears to be in agreement with the principle that intensive effort is secured in learning activity through creating a need. It is likely, however, that neither of these devices should be given prolonged use.

44. What types of learning exercises are most stimulating to learning activity in arithmetic?

While the experimental evidence is not of acceptable dependability, it seems reasonable that superior pupils are most stimulated by problems whose difficulty challenges them, and inferior pupils are most stimulated by problems which do not appear too difficult. In order that well motivated learning activity may be secured, problems set for the superior pupils may be stated in abstract terminology, may contain irrelevant elements, and may often be of the "puzzle" type, while those set for the inferior pupils will need to be in relatively concrete and familiar terminology or of the purely computational type.

45. In stimulating learning activity in arithmetic, what is the effectiveness of informing pupils of definite goals to be achieved?

Informing pupils of definite goals is an effective method of stimulating learning activity in arithmetic. While the experimental evidence is faulty, the conclusion conforms to the findings of research on the same problem in other subject-matter.

46. In stimulating learning activity in arithmetic, what is the effectiveness of informing pupils with respect to their status or progress?

The conclusions in favor of this method of stimulating learning activity are supported by dependable experimental evidence both from arithmetic and from other subject-matter.

47. What is the value of competition as a means of stimulating learning activity in arithmetic?

It is reported on the basis of fairly dependable experimental evidence that competition, and, particularly, individual rather than group competition, is an effective means of securing intensive and persistent effort. It is probable that excessive use of competition is undesirable, since it substitutes other objectives for those recognized as educational. There are occasions, however, when the use of competition is an effective device to bring a class of pupils out of a slump in learning by relieving the monotony of ordinary learning exercises.

48. What are the relative merits of commendation and reproof in stimulating learning activity in arithmetic?

Commendation and reproof are both reported to be stimulating to learning activity in arithmetic, but for most individuals commendation is a more beneficial stimulus than reproof. This conclusion is supported by reasonably dependable experimental evidence from the field of arithmetic and is in harmony with research on motivation in other subject-matter fields and in the psychological laboratory. The conclusion also conforms to the Law of Effect.

The contributions of research to the teaching of arithmetic. What constitutes a contribution depends upon the interpretation given to that term. It may be considered a contribution to show that an instructional procedure as applied to a particular group of pupils produces as satisfactory results, or nearly as satisfactory results, as another procedure may produce. Usually, however, a contribution is interpreted to mean the demonstration of the relative merits of two or more comparable procedures not merely for a particular group of pupils, but for all groups of pupils of a certain intellectual and educational status. If this more restricted interpretation is applied to the conclusions indicated in the preceding list, it is apparent that the dependable contributions of research in the teaching of arithmetic are relatively meager.

Probably the most significant contributions relate to the specific-

ity of calculation abilities and to the use of practice materials constructed so that adequate exercise is provided for each specific ability involved. Although research has not yet produced a complete and dependable list of the specific abilities in the field of arithmetical calculation, there are tentative lists for certain segments of this field, which appear to be rather highly dependable with reference to many of the items. The superiority of practice materials which provide for the exercise of each specific ability in proportion to the difficulty of attaining it has been demonstrated. It is, of course, not unlikely that, as these tentative lists of specific abilities are refined, superior practice materials may be devised, but this possibility does not detract from the fact that research has already contributed to the improvement of practice materials.

Closely related to this contribution is the demonstration of the effectiveness of diagnosis and of remedial instruction, and of systematic practice.

Research has contributed to an understanding of the nature of pupil responses to verbal problems and of the effect of introducing certain changes in the problem statement. Pupil responses to verbal problems are more satisfactory when they are stated in familiar terminology, and it appears that very little reasoning enters into the response of most pupils. Reading ability appears to be an important factor in the ability to respond to verbal problems, but the precise nature of its function has not been ascertained. Systematic training in finding the data given in a problem, in deciding upon calculations to be made, and in estimating the answer in round numbers is an effective procedure for teaching pupils to solve verbal problems.

Informing pupils of the status of their achievements in arithmetic is an effective means of securing intensity and persistence of effort in attaining higher levels of achievement. This procedure encourages each pupil to compete with his own past record. Competition between individual pupils and between groups is also effective.

There is considerable evidence that there is little or possibly no difference in the relative merits of several alternative calculation techniques. For example, the data secured in the studies of downward versus upward addition have been interpreted as favoring the latter technique, but the fact that the differences in achievement are so small that their significance is doubtful suggests the generalization just stated. This conclusion is also supported by *a priori* reasoning. If there is any significant difference in the relative merits of such alternative techniques it is likely that they would not be very apparent except on the higher levels of achievement, and since the function

of the school is not to produce highly expert calculators, it seems that the generalization stated at the beginning of this paragraph is the most significant contribution of the research attempting to evaluate alternative calculation techniques. Of course this generalization does not apply to cases in which one of the techniques is obviously time consuming or otherwise inefficient. For example, it should not be applied in support of "counting on the fingers."

Suggestions for research in the field of instructional methods in arithmetic. The evaluation and summary of research relating to the teaching of arithmetic afford a basis for some suggestions for future studies in this field. Although it is difficult to cite much definite evidence, the present writers have been impressed with the need for additional studies of verbal problems and of the nature of pupil responses to them. In the field of arithmetical calculation investigators have gone far in identifying the types of examples and the abilities involved in responding to them. It seems reasonable to assume that these are types of verbal problems. Research is needed to identify these types, if they exist. There is also need for more information about the function of reading in pupil responses to verbal problems and the relation of the form and vocabulary of problem statements to these responses.

Another suggested field of research relates to the instructional procedures employed in teaching pupils to solve problems. Should a method of analysis be employed? Should a complex problem be broken up into a series of simpler problems? Should a pupil be directed to compare the problem with ones he has solved and with solutions given in the text? What sort of attention should be given to the vocabulary? What types of learning exercises should be used in connection with verbal problems? Should pupils be taught a variety of problem types simultaneously or should each type be taught separately? To what extent and for what pupils is problem-solving activity stimulated by an occasional problem of the puzzle type? To what extent is the level of intelligence of the pupils a factor in generalization from number combinations specifically taught to those not taught? To what extent are flash cards used for drill purposes likely to engender improper eye-movement habits with respect to arithmetical subject-matter?

The possibility of evaluating comparable instructional procedures. The relatively meager contribution of the research summarized in this bulletin probably has suggested to the thoughtful reader the possibility that comparable instructional procedures cannot be evaluated with a high degree of precision. The evaluation of a procedure

by experimentation is dependent upon the control of all factors affecting the learning of pupils except the one being studied. The zeal and skill of the teacher in applying a given procedure affect the achievements of the pupils and these factors are difficult or impossible to control in many cases. Consequently it does not appear that precise and highly dependable evaluations of comparable instructional procedures should be expected. Attempts to determine the relative merit of certain "methods of teaching" will show that the procedures are approximately equal in merit, except when one of the procedures is distinctly inferior. In most such cases it is likely that a competent person could accurately predict this inferiority.

In support of this judgment, the requirements for precise and dependable evaluation of instructional procedures are briefly described.

REQUIREMENTS FOR PRECISE EVALUATION OF INSTRUCTIONAL METHODS IN ARITHMETIC²

1. Equivalent groups. The groups of pupils used in the experiment should be equivalent in all respects that will affect their arithmetical achievement during the experiment. This requirement can be approximated by pairing pupils on the basis of intelligence test scores and then comparing the groups thus formed with respect to chronological age, to previous achievement in the school subject, and to measures of arithmetical reading ability. If the differences between the means and the standard deviations of the groups with respect to these three characteristics are relatively small, the groups may be considered approximately equivalent. It is desirable that the groups also be approximately equivalent with respect to personality traits, physical conditions, sex, and race.

Two other techniques of securing equivalent groups may be suggested. The first is particularly adequate for investigations of the relative effectiveness of differing types of learning exercises. It is that of using such large groups that equivalence with respect to many factors is secured as a result of the operation of chance.³ It should be noted that this procedure is only feasible where the learning activity of the pupils is wholly directed by means of printed or mimeographed instructions. When this procedure is used the different groups are equally represented in all the classes participating in the experiment.

²These requirements have been taken with considerable adaptation from Monroe, W. S. and Engelhart, M. D. "Experimental Research in Education," *University of Illinois Bulletin*, Vol. 27, No. 32, Bureau of Educational Research Bulletin No. 48. Urbana: University of Illinois, 1930, p. 77-79.

³For a description of this technique, see: Monroe, W. S. "How Pupils Solve Problems in Arithmetic," *University of Illinois Bulletin*, Vol. 26, No. 23, Bureau of Educational Research Bulletin No. 44. Urbana: University of Illinois, 1929, p. 9. (79)

The second procedure which may be suggested is that used by Olander.⁴ This investigator paired pupils chiefly on the basis of growth in arithmetical ability over a period of five weeks during which the pupils were subjected to the same, or similar, instruction. The argument presented for this technique may be quoted here:

If two groups exhibit similar learning curves under similar instruction until a certain point is reached, it can be assumed that the groups are equal in the function in question. If a variation in the instruction of one group is then introduced which causes the learning curve of that group to rise abnormally, whereas the curve of the group under the unchanged technique continues to rise normally, it may be assumed that a difference in scores at any later point on the curve is attributable to the entrance of the variation in instruction.

2. Specification of experimental factor and control of non-experimental factors. The experimental factor should, if possible, be restricted to a single phase or detail of instructional procedure. The method used with the experimental group should vary from that used with the control group in only this single phase, and if other variations are permitted, their effect must be accurately measured or a plan of neutralization must be devised.⁵ The total instructional procedure to be used in both groups should be specified in writing, or at least a detailed record should be kept of what is done.

Controlled experimentation involves maintaining equal status for all factors in both the experimental and the control groups, except the single phase or detail of procedure which constitutes the experimental factor; or if the equal status is not maintained, the non-equivalence must be recognized and its effect on the experimental learning must be determined. The teacher factors whose control in arithmetic experiments appears to be the most important are (1) instructional techniques employed during the recitation period, especially those relating to the assignment, and motivation; (2) skill of the teacher in carrying out instructional techniques and classroom-management procedures; (3) zeal of the teacher; (4) personality traits of the teacher. In addition, care should be exercised to avoid marked differences in the minor teacher factors—physical condition, sex, and age.

The important factors under the head of general and extra-school factors are (1) materials of instruction, (2) environment in which learning activity takes place, and (3) minutes per day devoted to learning activity in arithmetic. The materials of instruction, desks, chairs, light, heat, ventilation, and other aspects of the learning environment should be identical for both groups. Study and recitation periods

⁴Olander, H. T. "Transfer of Learning in Simple Addition and Subtraction," *Elementary School Journal*, 31:363, January, 1931. (94)

⁵This requirement is sometimes designated as the Law of the Single Variable.

should be of equal length in the experimental and control group. Parents should be urged to refrain from influencing the arithmetical learning activity of the pupils, and, possibly, should be asked to cooperate in restricting the arithmetic learning activity to the classroom.

It should be noted that the precise prescription of an instructional procedure and the strict control of non-experimental factors is incompatible with good teaching. A teacher should adapt her techniques to the needs of her pupils as they become apparent. Hence conformity to the requirement for precise experimentation will, in many cases, tend to reduce the effectiveness of the teaching, and this in turn will introduce an element of uncertainty in the interpretation of the results of the experiment.

3. The measurement of achievement. In the consideration of the requirements under this head, the meaning of the validity of a test should be given careful attention. The problem of an experiment, when fully defined, either specifies or definitely implies the achievement to be measured. This achievement may be restricted to certain calculation skills or it may include also certain items of knowledge and certain general patterns of conduct. It may be restricted to the degree of ability possessed at the close of the period of experimentation, or it may consist of the residue after a period during which there is limited exercise of the ability.

A test that is highly valid for one purpose may be distinctly lacking in validity when used for another purpose. Consequently the validity of a test is a relative rather than an absolute characteristic, and this quality of one used in an experimental investigation can be determined only with reference to the specifications or implications of the problem. This means that the experimenter must assume the responsibility for determining the validity of the tests that he uses.

The reliability of a test refers to the variable errors in the resulting scores, assuming perfect validity. If the validity is also considered, any variable errors introduced because the achievement measured is not identical with that specified by the problem must be added to the effects of unreliability. Consequently the actual variable errors in the measures of achievement may be considerably larger than is indicated by the coefficient of reliability.

Finally the measures of achievement may involve constant or systematic errors.

4. The interpretation of differences in mean gains in achievement. In a typical experiment the treatment of the data results in a differ-

ence between the mean gains in achievement, or between the means of the final-test scores, of the experimental group and of the control group. If the groups are perfectly equivalent, if all non-experimental factors have been completely controlled, and if the measures of achievement are perfect—i.e., do not involve any errors, either variable or systematic—the obtained difference may be accepted as the actual difference in the mean gains of the two groups. These conditions are seldom, if ever, completely realized. Furthermore, when interpreting a difference in mean gains, the investigator usually desires to generalize—i.e., to make a statement with reference to the probability that the obtained difference has the same sign as the difference which might be obtained from any repetition of the experiment. The investigator may also wish to make a statement with reference to the probability that the obtained difference, in addition to having the same sign, is of the same order of magnitude as the difference which might be obtained from any repetition of the experiment. Hence, it is necessary to consider also the effect of sampling upon the data secured. In the following paragraphs attention is first directed to the statistical procedures to be employed in making allowances for variable errors of measurement and of sampling.

The statistical procedures outlined in the following paragraphs yield the standard⁶ error of the difference in mean gains, or of the difference between final-test means, due to the combined⁷ effect of variable errors of measurement and variable errors of sampling. If the difference in mean gains, or final-test means, is equal to, or greater than, 2.78 times the standard error of the difference, or 4.4 times the probable error of the difference, it is customary to recognize the difference as "statistically" significant. The statement may be made in interpretation, that the chances are 369 to 1, or better, that the sign of the obtained difference is not due to the combined effect of the *variable errors of measurement and the variable errors of sampling*.

The chances that the true difference does not differ from the obtained difference by more than plus or minus the standard error of the difference are 2.15 to 1, by more than plus or minus twice the standard error of the difference, 21 to 1, and by more than plus or minus three times the standard error of the difference, 369 to 1. This interpretation may be used when the investigator is interested in stating the

⁶The probable error may be obtained by multiplying the standard error by the constant, .6745.

⁷For a discussion of the fact that $\frac{\sigma}{\sqrt{N}}$ allows for the combined effect of variable errors of measurement and of sampling, see:

Kelley, T. L. "Note upon Holzinger's Formula for the Probable Error," *Journal of Educational Psychology*, 14:376-77, September, 1923.

Huffaker, C. L. and Douglass, H. R. "On the Standard Errors of the Mean Due to Sampling and to Measurement," *Journal of Educational Psychology*, 19:643-49, December, 1928.

probabilities that the true difference is of the same order of magnitude as well as of the same sign as the obtained difference.⁸

The maximum allowance which needs to be made for the combined effect of variable errors of measurement and variable errors of sampling may be determined by means of the following formulae in which σ_e and σ_c are the standard deviations of the distributions of individual gains of the experimental and of the control pupils:

$$\sigma_{\text{Mean Gain E}} = \frac{\sigma_e}{\sqrt{N}}$$

$$\sigma_{\text{Mean Gain C}} = \frac{\sigma_c}{\sqrt{N}}$$

$$\sigma_{\text{Difference}} = \sqrt{\sigma_{\text{Mean Gain E}}^2 + \sigma_{\text{Mean Gain C}}^2}$$

If equivalent forms of an arithmetic test are not used at the beginning and end of the experiment, or if scores are not converted into comparable units, calculation of individual gains is impossible.⁹

The subtraction of a pupil's initial-test score from his final-test score is justified only when the scores are in terms of approximately equal units—a condition approached when equivalent forms of a test are used, or when scores are converted into comparable units. When equivalent forms are not used or conversion has not been resorted to, comparison is restricted to the difference between the final-test means. In this case the standard deviations, σ_e and σ_c , refer to the distributions of final-test scores of the experimental and of the control pupils. The first two formulae will then yield the standard errors of the final-test means, and the third formula, when the squares of the standard errors of the final-test means are inserted under the radical, will yield the standard error of the difference between the final-test means.

It was stated in introducing the formulae given above that they provide the maximum allowance which needs to be made for the combined effect of variable errors of measurement and variable errors of sampling. Two reasons may be given in support of this statement.

⁸For a table of these probabilities, see:

Monroe and Engelhart, *op. cit.*, p. 66.

⁹If the pupils start the experiment with zero arithmetic ability, the scores on the final test represent gains. If the tests used at the beginning and end of the experiment are equally valid measures of the experimental achievement, although not equivalent forms, conversion of the initial and final measures into standard scores, T-scores, or grade scores makes possible the calculation of individual gains.

The first reason is that $\frac{\sigma}{\sqrt{N}}$, in addition to measuring the effect of variable errors of measurement, measures the effect of chance where the operation of chance in the selection of the groups is not restricted. In the following paragraphs it is indicated that the procedure usually employed in securing equivalent groups—i.e., pairing pupils with respect to intelligence test scores, or making adjustments so that means and standard deviations of the two groups are equal even though pupils are not paired “pupil for pupil”—tends to reduce the effect of chance. The formulae given above yield a precise allowance for the effect of chance in the selection of the groups, only where the groups are both random with respect to the population from which they were drawn and with respect to each other.

The second reason for stating that the formulae given above provide a maximum allowance is that these formulae neglect the correlation that may exist between the gains of the *paired* pupils, or between their final-test scores. In other words, the expression, $-2r_{ge\ ge} \sigma_{\text{Mean Gain E}} \cdot \sigma_{\text{Mean Gain C}}$, where $r_{ge\ ge}$ is the coefficient obtained by correlating the distribution of individual gains of the experimental pupils with the distribution of individual gains of the control pupils, should also be included under the radical of the third formula given above. Coefficients of correlation are regularly obtained by correlating two distributions of measures of the same individuals. The uncertain conclusions of research on the effect of practice on individual differences would cause one to question the dependability of a coefficient obtained by correlating gains of paired individuals. Owing to the uncertainty of this correlation and owing to the reduction in the operation of chance where procedures are employed to secure equivalence, the standard errors obtained through the utilization of the formulae given above should be interpreted as *limits beyond which the true standard errors cannot fall*.

Lindquist has stated with regard to the formulae given above that they are based on “the assumption that the samples used are *strictly random* selections from the populations they represent.”¹⁰ He continues:

This assumption is not applicable to matched groups. The process of matching on the basis of a measure which is correlated with the final measure *destroys* the randomness of the samples with respect to this final measure. The probable amount of sampling error in the obtained difference, instead of being as large as that indicated

¹⁰Lindquist, E. F. “The Significance of a Difference Between ‘Matched’ Groups,” *Journal of Educational Psychology*, 22:198, March, 1931.

Advice with respect to the formulae given on page 104 received through correspondence with Dr. Lindquist is deeply appreciated.

by the formulas given above, is usually considerably less, in some cases by more than fifty percent.¹¹

The allowance to be made for the combined effect of variable errors of measurement and variable errors of sampling in the case of paired, or matched, groups may be determined by means of the following formulae:¹²

$$\sigma_{\text{Mean Gain E}} = \frac{\sigma_e \sqrt{1 - r_{ie}^2}}{\sqrt{N}}$$

$$\sigma_{\text{Mean Gain C}} = \frac{\sigma_c \sqrt{1 - r_{ic}^2}}{\sqrt{N}}$$

$$\begin{aligned} \sigma_{\text{Difference}} \\ \text{Mean Gain E} - \text{Mean Gain C} \end{aligned} = \sqrt{\sigma_{\text{Mean Gain E}}^2 + \sigma_{\text{Mean Gain C}}^2}$$

In the formulae given above σ_e and σ_c are the standard deviations of the distributions of individual gains of the pupils in the experimental and in the control groups. The coefficient of correlation, r_{ie} , refers to the relation between the intelligence test scores, or other measures, used in pairing the experimental pupils and their corresponding individual gains.¹³ The coefficient of correlation, r_{ic} , refers to the relation between the intelligence test scores, or other measures used in pairing the control pupils, and their corresponding individual gains.¹⁴ Lindquist suggests that where the methods compared are unlikely to result in producing a significant difference in r_{ie} and r_{ic} , the statistical technique may be simplified by using the formula:¹⁵

$$\begin{aligned} \sigma_{\text{Difference}} \\ \text{Mean Gain E} - \text{Mean Gain C} \end{aligned} = \sqrt{\left(\frac{\sigma_e^2}{N_e} + \frac{\sigma_c^2}{N_c} \right) (1 - r^2)}$$

¹¹Lindquist, *op. cit.*, p. 198.

¹²For rigorous mathematical proof, see:

Wilks, S. S., "The Standard Error of the Means of Matched Samples," *Journal of Educational Psychology*, 22:205-08, March, 1931.

¹³The scores used in pairing may be intelligence test scores, or they may be the scores of the pupils of the experimental group on the form of the arithmetic test administered at the beginning of the experiment. If the pupils are paired on the basis of measures of two, or more, traits, the coefficient of multiple correlation, expressing the relationship between the individual gains and these traits in combination, is the appropriate coefficient to use. See:

Wilks, *op. cit.*, p. 208.

¹⁴Where equivalent forms of the same subject-matter test are not administered at the beginning and the end of the experiment, or where scores have not been converted into comparable units, calculation of individual gains is impossible. In using the above formulae, σ_e and σ_c should represent the standard deviations of the distributions of final-test scores of the experimental and the control groups, and r_{ie} and r_{ic} should represent the relationships between the test scores used in pairing and the final-test scores of the experimental and control pupils respectively. The third formula then yields the standard error of the difference between final-test means for matched groups.

¹⁵Lindquist, *op. cit.*, p. 202-03.

The formula given by Lindquist has been slightly modified by the authors to represent the standard error of the difference in mean gains, and the symbols have been changed.

N_e and N_c represent the numbers of pupils in the experimental and control groups, usually the same, σ_e and σ_c , the standard deviations of the two distributions of individual gains, and r stands for the relationship existing between the measures of *all* the pupils used in pairing and their individual gains.

It should be noted in this connection that Lindquist suggests that the formulae given above "should be valid for use with groups that have not been matched 'pupil for pupil', but in which the means and standard deviations alone have been equated." He adds, however, that "a more rigid mathematical proof of this proposition should be provided before much confidence is placed in it."¹⁶ The following quotation is indicative of just what is allowed for when these formulae are used:

It is also important to note that formula (9) (the one just given) does not indicate how far the obtained difference between two matched samples is likely to deviate from the difference that would have been obtained had the *entire* population been measured, but tells only how far the obtained difference is likely to deviate from the difference that would have been found between infinitely large groups showing the *same distribution of initial measures as that of the matched samples that were used*.¹⁷

It will be seen from the statement quoted above that generalizations in which this standard error of difference is used apply to "infinitely large groups showing the same distribution of initial measures as that of the matched samples." If the matched samples are, for example, somewhat superior in intelligence to the average intelligence of the general population from which the samples were drawn, strictly speaking, the generalizations apply to similar matched samples. When N is greater than 30, the experimental group is selected in a random fashion from the general population; and the control group is obtained by selecting pupils from the general population who match the experimental pupils, this standard error of difference may be used with considerable justification in formulating conclusions relative to the general population.¹⁸

Finally, it should be noted that the standard error of difference obtained by the formulae suggested by Lindquist neglects the correlation that may exist between the gains of the paired pupils, or their final-test scores. Hence, the standard error of difference so obtained is also to be interpreted as a limit beyond which the true standard error cannot fall. The limit, however, is probably closer to the true

¹⁶Lindquist, *op. cit.*, p. 202.

¹⁷*Ibid.*, p. 202-03.

Literally speaking, "infinitely large groups" would include the "entire population" and thus have the same distributions of initial measures. The term "infinitely large groups" as used here refers to groups so large that the effects of variable errors are altogether negligible, but groups which are not sufficiently large in size to include the entire population. "Infinitely large groups" of this type may have distributions differing from those of the "entire population."

¹⁸*Ibid.*, p. 203.

standard error than that obtained when $\frac{\sigma}{\sqrt{N}}$ is used in computing the standard error of the mean gains, or the final-test means.

When the groups of pupils are ordinary school classes, and not random samples, statistical procedure may be employed in making allowance for variable errors of measurement alone. In preceding paragraphs it was indicated that these errors are of minor importance in comparison with the other possible sources of undependability of a difference in mean gains, or in final-test means. Where reasonably reliable tests have been used, and the experimental and control groups are fairly large, the computation of the standard error of measurement of the difference in mean gains, or final-test means, is unlikely to contribute much to the meaning of the findings. If the difference in mean gains is comparatively large the investigator is justified in assuming that the dependability of the difference, *so far as the groups used in the experiment are concerned*, is not significantly affected by variable errors of measurement.

The procedures just described constitute a means for calculating the probable effect upon the difference of the mean gains, or final-test means, of only the combined effect of the variable errors of measurement and of the error of sampling. Unfortunately it is not possible to calculate the probable effects of the systematic errors of measurement in either, or both, the first and second trial scores, the invalidity of the test, as determined by the problem of the experiment, and any lack of control of significant non-experimental factors. In general the experimenter can only estimate the probable effects of these conditions. Usually some circumstantial evidence can be cited in support of his estimate, but the uncertainty of any estimate justifies the assertion that the determination of "statistical" significance by means of the formulae given above should not be treated very seriously. The interpretation of a small difference in mean gains, or final-test means, will usually be uncertain even when it is shown to be "statistically" significant.

The statement just made is important. The use of statistical formulae, especially when somewhat complex, tends to be impressive, and when the difference is shown to be "statistically" significant there is doubtless a suggestion to the uninformed that all limitations of the data have been allowed for. This is not the case. As a matter of fact it is reasonably apparent that, in many cases, the probable error of the difference in mean gains is an index of the least significant limitations of the data. By way of emphasizing this point it may be suggested that when attempting to interpret a difference an

experimenter should focus his attention upon the limitations of the data whose probable effect cannot be calculated by a formula.

5. Generalization. Consideration of the probable effect of the variable errors of sampling has, of course, related to the generalizing of the data secured in a particular experiment. The formulae presented furnish a statistical basis for generalizing only when both the control group and the experimental group constitute random samples from the larger population, or when one of the two equated groups has been selected at random.

Generalization appears justified, where random sampling has not been employed, when it can be shown that *lack of representativeness of the groups* does not seriously limit the dependability of the difference in achievement in favor of a given method. In showing that the groups used in the experiment are sufficiently typical or representative to justify generalization, the experimenter should present all available evidence relative to the traits of the groups concerned. For example, the intelligence test scores will be known, and the experimenter should show how the mean and the standard deviation of these scores compare with corresponding measures of the larger population. If the available evidence indicates that the groups are highly representative of the larger population, he may generalize with considerable confidence; if the evidence indicates that the groups are not reasonably representative of the larger population, he must refrain from generalizing or appropriately limit his statements.

Feasibility versus evaluation (effectiveness) of instructional techniques. In closing this discussion it seems appropriate to comment upon the demonstration of the feasibility of a procedure versus the determination of the relative merits of two or more specified procedures. The former can be accomplished by a single-group experiment. It is only necessary to show that as applied by a certain teacher or group of teachers the procedure resulted in reasonably satisfactory achievements. To determine the relative merits of two or more specified procedures controlled experimentation is required. The difficulties encountered in controlling non-experimental factors and in securing accurate and valid measures of the achievement specified by the problem of the experiment have been noted in the preceding pages. It is apparent that the expectation of precise evaluation is not justified.

CHAPTER IX

BIBLIOGRAPHY OF INVESTIGATIONS PERTAINING TO METHODS OF TEACHING ARITHMETIC

1. Adams, R. E. "A Study of the Comparative Value of Two Methods of Improving Problem Solving Ability in Arithmetic," Philadelphia: University of Pennsylvania, 1930. 68 p.
2. Anspaugh, G. E. "Teaching the Number Facts in the Komensky School," Chicago Principals' Club, Second Yearbook. Chicago: Chicago Principals' Club, 1927, p. 88-89.
3. Anthony, Kate, et al. "The Development of Proper Attitudes Towards School Work," *School and Society*, 2:926-34, December 25, 1915.
4. Arnett, L. D. "Counting and Adding," *American Journal of Psychology*, 16:327-36, July, 1905.
5. Ballard, P. B. "Norms of Performance in the Fundamental Processes of Arithmetic, with Suggestions for Their Improvement," *Journal of Experimental Pedagogy*, 2:396-405, December 5, 1914; 3:9-20, March 5, 1915.
6. Ballenger, H. L. "Overcoming Some Addition Difficulties," *Journal of Educational Research*, 13:111-17, February, 1926.
7. Ballou, F. W. "Improving Instruction through Educational Measurement," *Educational Administration and Supervision*, 2:354-67, June, 1916.
8. Beatty, W. W. "The Additive versus the Borrowing Method of Subtraction," *Elementary School Journal*, 21:198-200, November, 1920.
9. Beito, E. A. and Brueckner, L. J. "A Measurement of Transfer in the Learning of Number Combinations," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 569-87.
10. Bowman, H. L. "The Relation of Reported Preference to Performance in Problem Solving," *University of Missouri Bulletin*, Vol. 30, No. 36, Education Series, No. 29. Columbia: University of Missouri, 1929. 52 p.
11. Bradford, E. J. G. "Suggestion, Reasoning, and Arithmetic," *Forum of Education*, 3:3-12, February, 1925.
12. Brooks, S. S. "A Study of the Technical and Semi-technical Vocabulary of Arithmetic," *Educational Research Bulletin (Ohio State University)*, 5:219-22, May 26, 1926.
13. Brown, J. C. "An Investigation on the Value of Drill Work in the Fundamental Operations of Arithmetic," *Journal of Educational Psychology*, 2:81-88, February, 1911; 3:485-92, 561-70; November, December, 1912.
14. Brownell, W. A. "Remedial Cases in Arithmetic," *Peabody Journal of Education*, 7:100-107, September, 1929.
15. Buckingham, B. R. "Adding Up or Down: A Discussion," *Journal of Educational Research*, 12:251-61, November, 1925.
16. Buckingham, B. R. "The Additive versus the Take-away Method of Teaching the Subtraction Facts," *Educational Research Bulletin (Ohio State University)*, 6:265-69, September 28, 1927.

17. **Buckingham, B. R.** "Teaching Addition and Subtraction Facts Together or Separately," *Educational Research Bulletin* (Ohio State University) 6:228-29, 240-42; May 25, 1927.
18. **Buckingham, B. R.** "Upward versus Downward Addition," *Journal of Educational Research*, 16:315-22, December, 1927.
19. **Burton, C. B.** "Results of Definite Drill in the Four Fundamental Processes as Shown by the Woody-McCall Mixed Fundamentals," *Fifth Year-book of the Department of Elementary School Principals*. Washington: National Education Association, 1926, p. 323-28.
20. **Buswell, G. T. and John, Lenore.** "Diagnostic Studies in Arithmetic," *Supplementary Educational Monographs*, No. 30. Chicago: University of Chicago Press, 1926. 212 p.
21. **Buswell, G. T. and John, Lenore.** "The Vocabulary of Arithmetic," *Supplementary Educational Monographs*, No. 38. Chicago: University of Chicago Press, 1931. 146 p.
22. **Chapman, J. C. and Feder, R. B.** "The Effect of External Incentives on Improvement," *Journal of Educational Psychology*, 8:469-74, October, 1917.
23. **Chase, S. E.** "Waste in Arithmetic," *Teachers College Record*, 18:360-70, September, 1917.
24. **Chase, V. E.** "The Diagnosis and Treatment of Some Common Difficulties in Solving Arithmetic Problems," *Journal of Educational Research*, 20:335-42, December, 1929.
25. **Clapp, F. L., Chase, W. J., and Merriman, Curtis.** "A Study of the Effectiveness of Two Kinds of Teaching Material," *Introduction to Education*. Boston: Ginn and Company, 1929, p. 420-24.
26. **Clark, J. R. and Vincent, E. L.** "A Comparison of Two Methods of Arithmetic Problem Analysis," *Mathematics Teacher*, 18:226-33, April, 1925.
27. **Clark, J. R. and Vincent, E. L.** "A Study of the Effect of Checking upon Accuracy in Addition," *Mathematics Teacher*, 19:65-71, February, 1926.
28. **Clemens, P. B. and Neubauer, P. F.** "A Supervision Project in Multiplication," *Journal of Educational Research*, 18:387-96, December, 1928.
29. **Cole, L. W.** "Adding Upward and Downward," *Journal of Educational Psychology*, 3:83-94, February, 1912.
30. **Collier, Myrtle.** "Learning to Multiply Fractions," *School Science and Mathematics*, 22:324-29, April, 1922.
31. **Collings, Ellsworth.** *An Experiment with a Project Curriculum*. New York: Macmillan Company, 1923. 346 p.
32. **Conard, H. E. and Arps, G. F.** "An Experimental Study of Economical Learning," *American Journal of Psychology*, 27:507-29, October, 1916.
33. **Courtis, S. A.** "The Courtis Standard Tests in Boston, 1912-1915: An Appraisal," *School Document No. 15*, 1916. Department of Educational Investigation and Measurement Bulletin No. 10. Boston: Boston Printing Department, 1916. 48 p.
34. **Dawson, C. D.** "Some Results in Using Starch's Arithmetic Reasoning Test," *Journal of Educational Research*, 2:677-78, October, 1920.

35. Drushel, J. A. "A Study of the Amount of Arithmetic at the Command of High-School Graduates Who Have Had No Arithmetic in Their High-School Course," *Elementary School Journal*, 17:657-61, May, 1917.
36. Edaña, Tiburcio. "An Analysis of Arithmetic Textual Matter," *Philippine Public Schools*, 1:81-84, February, 1928.
37. Evans, J. E. and Knoche, F. E. "The Effects of Special Drill in Arithmetic as Measured by the Woody and the Courtis Arithmetic Tests," *Journal of Educational Psychology*, 10:263-76, May-June, 1919.
38. Fowlkes, J. G. "A Report of a Controlled Study of the Learning of Multiplication by Third-Grade Children," *Journal of Educational Research*, 15:181-89, March, 1927.
39. Gabbert, M. L. "Remedial Cases in Arithmetic; Case 1," *Peabody Journal of Education*, 7:147-55, November, 1929.
40. Gates, A. I., Batchelder, M. I., and Betzner, Jean. "A Modern Systematic versus an Opportunistic Method of Teaching," *Teachers College Record*, 27:679-700, April, 1926.
41. Graham, V. O. "Arithmetic Reasoning Project and the Measurement of Improvement," *Chicago Principals' Club, Second Yearbook*. Chicago: Chicago Principals' Club, 1927, p. 86-87.
42. Greene, H. A. "Directed Drill in the Comprehension of Verbal Problems in Arithmetic," *Journal of Educational Research*, 11:33-40, January, 1925.
43. Guiler, W. S. "Improving Computational Ability," *Elementary School Journal*, 30:111-16, October, 1929.
44. Hackler, J. M. "The Relation between Successful Progress in Mathematics and the Ability to Read and Understand, and the Factors That Contribute to Success or Failure in Mathematics." Unpublished master's thesis in Education. Chicago: University of Chicago, 1921. 82 p.
45. Hagen, H. H. "A Study of Practice Periods in Arithmetic Fundamentals," *Chicago Principals' Club, Second Yearbook*. Chicago: Chicago Principals' Club, 1927, p. 93-95.
46. Hahn, H. H. and Thorndike, E. L. "Some Results of Practice in Addition under School Conditions," *Journal of Educational Psychology*, 5:65-84, February, 1914.
47. Hanna, P. R. "Arithmetic Problem Solving: A Study of the Relative Effectiveness of Three Methods of Problem Solving." New York: Bureau of Publications, Teachers College, Columbia University, 1929. 68 p.
48. Harlan, C. L. "Years in School and Achievements in Reading and Arithmetic," *Journal of Educational Research*, 8:145-49, September, 1923.
49. Hurlock, E. B. "An Evaluation of Certain Incentives Used in School Work," *Journal of Educational Psychology*, 16:145-59, March, 1925.
50. Hyde, L. L. and Clapp, F. L. "Elements of Difficulty in the Interpretation of Concrete Problems in Arithmetic," *Bureau of Educational Research Bulletin No. 9*. Madison: University of Wisconsin, 1927. 84 p.
51. John, Lenore. "Difficulties in Solving Problems in Arithmetic," *Elementary School Journal*, 31:202-15, November, 1930.
52. Johnston, J. T. "The Merits of Different Methods of Subtraction," *Journal of Educational Research*, 10:279-90, November, 1924.

53. Kallom, A. W. "The Importance of Diagnosis in Educational Measurement," *Journal of Educational Psychology*, 10:1-12, January, 1919.
54. Kathleen, Sister M. "Some Results of Remedial Instruction Following the Use of Diagnostic Arithmetic Tests," *Catholic Educational Review*, 24:19-27, January, 1926.
55. Kelly, F. J. "The Results of Three Types of Drill on the Fundamentals of Arithmetic," *Journal of Educational Research*, 2:693-700, November, 1920.
56. Kerr, M. A. "Effects of Six Weeks Daily Drill in Arithmetic," *Studies in Arithmetic*, Indiana University Studies No. 32. Bloomington: Indiana University, 1916, p. 79-95.
57. Kirby, T. J. "Practice in the Case of School Children," *Teachers College, Columbia University Contributions to Education*, No. 58. New York: Bureau of Publications, Teachers College, Columbia University, 1913. 98 p.
58. Kirkpatrick, E. A. "An Experiment in Memorizing versus Incidental Learning," *Journal of Educational Psychology*, 5:405-12, September, 1914.
59. Kline, L. W. and Anderson, P. K. "The Role of Habit in Reasoning," *School Science and Mathematics*, 26:156-67, February, 1926.
60. Knight, F. B. "The Superiority of Distributed Practice in Drill in Arithmetic," *Journal of Educational Research*, 15:157-65, March, 1927. (A summary of 61.)
61. Luse, E. M. "Transfer within Narrow Mental Functions. A Study of the Effects of Distributed versus Non-Distributed Drill in Arithmetic," *University of Iowa Monograph in Education* No. 5. Iowa City: University of Iowa, 1925.
62. Knight, F. B. and Setzafandt, A. O. H. "Transfer within a Narrow Mental Function," *Elementary School Journal*, 24:780-87, June, 1924.
63. Krause, A. K. "Why Monroe Diagnostic Tests in Arithmetic?" *Contributions to Education*, Vol. 2. Yonkers, New York: World Book Company, 1928, p. 15-17.
64. Kulp, C. L. "A Method of Securing Real-Life Problems in the Fundamentals of Arithmetic," *Elementary School Journal*, 29:428-30, February, 1929.
65. Kulp, C. L. "A Study of the Relative Effectiveness of Two Types of Standard Arithmetic Practice Materials," *Journal of Educational Research*, 22:381-387, December, 1930.
66. Lazar, May. *Diagnostic and Remedial Work in Arithmetic Fundamentals for Intermediate Grades*. New York: Board of Education, 1928. 204 p.
67. Lessenger, W. E. "Reading Difficulties in Arithmetical Computation," *Journal of Educational Research*, 11:287-91, April, 1925.
68. Lutes, O. S. "An Evaluation of Three Techniques for Improving Ability to Solve Arithmetic Problems: A Study in the Psychology of Problem Solving," *University of Iowa Monographs in Education*, Series 1, No. 6. Iowa City: University of Iowa, 1926. 42 p.
69. McClelland, W. W. "An Experimental Study of the Different Methods of Subtraction," *Journal of Experimental Pedagogy*, 4:293-99, December 5, 1918.
70. Maller, J. B. "Cooperation and Competition, An Experimental Study in Motivation," *Teachers College, Columbia University Contributions to*

- Education, No. 384. New York: Bureau of Publications, Teachers College, Columbia University, 1929. 176 p.
71. Mead, C. D. and Johnson, C. W. "Testing Practice Material in the Fundamentals," *Journal of Educational Psychology*, 9:287-97, May, 1918.
 72. Mead, C. D. and Sears, Isabel. "Additive Subtraction and Multiplicative Division Tested," *Journal of Educational Psychology*, 7:261-70, May, 1916.
 73. Meriam, J. L. "How Well May Pupils Be Prepared for High School Work without Studying Arithmetic, Grammar, etc., in the Grades?" *Journal of Educational Psychology*, 6:361-64, June, 1915.
 74. Merton, E. L., et al. "Remedial Work in Arithmetic." *Second Yearbook of the Department of Elementary School Principals*. Washington: National Education Association, 1923, p. 395-429.
 75. Messick, A. I. "Effect of Certain Types of Speed Drills in Arithmetic," *Mathematics Teacher*, 19:104-9, February, 1926.
 76. Mitchell, Claude. "The Specific Type of Problem in Arithmetic versus the General Type of Problem," *Elementary School Journal*, 29:594-96, April, 1929.
 77. Monroe, W. S. "The Ability to Place the Decimal Point in Division," *Elementary School Journal*, 18:287-93, December, 1917.
 78. Monroe, W. S. "The Derivation of Reasoning Tests in Arithmetic," *School and Society*, 8:295-99, 324-29; September 7, 14, 1918.
 79. Monroe, W. S. "How Pupils Solve Problems in Arithmetic," *University of Illinois Bulletin*, Vol. 26, No. 23, Bureau of Educational Research Bulletin No. 44. Urbana: University of Illinois, 1929. 31 p.
 80. Morgan, L. D. "Specific vs. General Drill in the Fundamentals of Arithmetic," *School Science and Mathematics*, 29:528-29, May, 1929.
 81. Morton, R. L. "An Analysis of Pupils' Errors in Fractions," *Journal of Educational Research*, 9:117-25, February, 1924.
 82. Morton, R. L. "An Analysis of Errors in the Solution of Arithmetic Problems," *Educational Research Bulletin (Ohio State University)*, 4:187-90, April 29, 1925.
 83. Myers, G. C. "Imagination in Arithmetic," *Journal of Education*, 105:662-63, June 13, 1927.
 84. Myers, G. C. "The Price of Speed Pressure in the Learning of Number," *Educational Research Bulletin (Ohio State University)*, 7:265-68, September 19, 1928.
 85. Myers, G. C. and Myers, C. E. "The Cost of Quick Shifting in Number Learning," *Educational Research Bulletin (Ohio State University)*, 7:327-34, October 31, 1928.
 86. Myers, G. C. and Myers, C. E. "Finding Mistakes versus Correct Associations in Simple Number-Learning," *Journal of Educational Research*, 18:25-31, June, 1928.
 87. Neal, E. A. and Foster, Inez. "An Experiment with Remedial Work in Common Fractions," *Elementary School Journal*, 29:280-83, December, 1928.
 88. Newcomb, R. S. "Effective Drill Exercises in Arithmetic," *Journal of Educational Psychology*, 16:127-31, February, 1925.

89. Newcomb, R. S. "Securing the Maximum Amount of Work from Every Pupil," *Elementary School Journal*, 25:376-79, January, 1925.
90. Newcomb, R. S. "Teaching Pupils How to Solve Problems in Arithmetic," *Elementary School Journal*, 23:183-89, November, 1922.
91. Norem, G. B. and Knight, F. B. "The Learning of the One Hundred Multiplication Combinations," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 551-68.
92. O'Brien, F. P. "Co-operative Experiment Pertaining to Instruction in Arithmetic," *American Education*, 31:219-21, February, 1928.
93. O'Brien, F. P. "Improvement of Instruction in Arithmetic," *University of Kansas Bulletin of Education*, Vol. 1, No. 4. Lawrence: University of Kansas, 1927. 42 p.
94. Olander, H. T. "Transfer of Learning in Simple Addition and Subtraction," *Elementary School Journal*, 31:358-69, 427-37; January, February, 1931.
95. Osburn, W. J. and Drennan, L. J. "Problem Solving in Arithmetic," *Educational Research Bulletin (Ohio State University)*, 10:123-28, March 4, 1931.
96. Otto, H. J. "Remedial Instruction in Arithmetic," *Elementary School Journal*, 28:124-33, October, 1927. Also in: *Journal of the National Education Association*, 17:87-89, March, 1928.
97. Overman, J. R. "An Experimental Study of the Effect of the Method of Instruction on Transfer of Training in Arithmetic," *Elementary School Journal*, 31:183-90, November, 1930.
98. Panlasigui, Isidoro and Knight, F. B. "The Effect of Awareness of Success or Failure," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 611-619.
99. Partridge, C. M. "Number Needs in Children's Reading Activities," *Elementary School Journal*, 26:356-66, January, 1926.
100. Phillips, F. M. "Value of Daily Drill in Arithmetic," *Journal of Educational Psychology*, 4:159-63, March, 1913.
101. Reavis, W. C. "The Social Motive in the Teaching of Arithmetic," *Elementary School Journal*, 18:264-67, December, 1917.
102. Reed, H. B. "Distributed Practice in Addition," *Journal of Educational Psychology*, 15:248-49, April, 1924.
103. Repp, A. C. "Mixed versus Isolated Drill Organization," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 535-49.
104. Richardson, J. W. "The Campaign Method in Elementary Education," *Journal of Educational Research*, 2:481-92, June, 1920.
105. Rosse, J. C. "An Experiment to Test the Increase in Reasoning Ability from the Use of Test and Practice Sheets in 6A Arithmetic," *Journal of Educational Research*, 22:210-13, October, 1930.
106. Sheerin, E. M. "Application of the Dalton Plan to Teaching Arithmetic," *Contributions to Education*, Vol. 2. Yonkers, New York: World Book Company, 1928, p. 18-22.

107. Smith, J. H. "Individual Variations in Arithmetic," *Elementary School Journal*, 17:195-200, November, 1916.
108. Soth, M. R. "A Study of a Pupil Retarded in Arithmetic," *Elementary School Journal*, 29:439-42, February, 1929.
109. Springer, Isadore "Teaching Denominate Numbers," *Journal of Educational Psychology*, 6:630-32, December, 1915.
110. Steinway, L. S. "An Experiment in Games Involving a Knowledge of Number," *Teachers College Record*, 19:43-53, January, 1918.
111. Stevenson, P. R. "Difficulties in Problem Solving," *Journal of Educational Research*, 11:95-103, February, 1925.
112. Stevenson, P. R. "Increasing the Ability of Pupils to Solve Arithmetic Problems," *Educational Research Bulletin (Ohio State University)*, 3:267-70, October 15, 1924.
113. Stone, C. W. "An Experimental Study in Improving Ability to Reason in Arithmetic," *Twenty-Ninth Yearbook of the National Society for the Study of Education*. Bloomington, Illinois: Public School Publishing Company, 1930, p. 589-99.
114. Taylor, J. S. "Subtraction by the Addition Process," *Elementary School Journal*, 20:203-7, November, 1919.
115. Terry, P. W. "How Numerals are Read: An Experimental Study of the Reading of Isolated Numerals and Numerals in Arithmetic Problems," *Supplementary Educational Monographs*, No. 18. Chicago: University of Chicago Press, 1922. 110 p. For a brief report see reference No. 116.
116. Terry, P. W. "The Reading Problem in Arithmetic," *Journal of Educational Psychology*, 12:365-77, October, 1921.
117. Washburne, C. W. "Comparison of Two Methods of Teaching Pupils to Apply the Mechanics of Arithmetic to the Solution of Problems," *Elementary School Journal*, 27:758-67, June, 1927.
118. Washburne, C. W. and Morphett, M. V. "Unfamiliar Situations as a Difficulty in Solving Arithmetic Problems," *Journal of Educational Research*, 18:220-24, October, 1928.
119. Washburne, C. W. and Osborne, Raymond. "Solving Arithmetic Problems," *Elementary School Journal*, 27:219-26, 296-304; November, December, 1926.
120. Wertheimer, J. E. "Some Results of Monroe's Diagnostic Tests in Arithmetic," *Journal of Educational Psychology*, 11:109-112, February, 1920.
121. Wheat, H. G. "The Relative Merits of Conventional and Imaginative Types of Problems in Arithmetic," *Teachers College, Columbia University Contributions to Education*, No. 359. New York: Bureau of Publications, Teachers College, Columbia University, 1929. 124 p.
122. Wilson, Estaline. "Improving the Ability to Read Arithmetic Problems," *Elementary School Journal*, 22:380-86, January, 1922.
123. Wilson, G. M. "New Standards in Arithmetic: A Controlled Experiment in Supervision," *Journal of Educational Research*, 22:351-60, December, 1930.

124. Wimmer, H. "Experimental Study of the Effects of Drill in Arithmetic Processes under Varying Conditions," *Indiana University Studies* No. 32. Bloomington: Indiana University, 1916, p. 96-102.
125. Winch, W. H. "'Equal Additions' versus 'Decomposition' in Teaching Subtraction: An Experimental Research," *Journal of Experimental Pedagogy*, 5:207-20, 261-70; June 5, December 6, 1920.
126. Winch, W. H. "Should Young Children Be Taught Arithmetical Proportion?" *Journal of Experimental Pedagogy*, 2:79-88, 319-30, 406-20; June, 1913; June 5, December 5, 1914; 3:89-95, June 5, 1915.
127. Yeager, William. "Diagnosis and Remedial Measures in Arithmetic," *Twelfth Annual Schoolmen's Week Proceedings*. Philadelphia: University of Pennsylvania, 1925, p. 296-305.
128. "Methods of Subtraction," *St. Louis Public School Messenger*, 26:28-32, September 1, 1928.

